How to use the Kohonen algorithm for forecasting

Marie Cottrell

SAMOS-MATISSE, Université Paris 1

(with Bernard Girard, Patrice Gaubert, Patrick Letrémy, Patrick Rousset, Joseph Rynkiewicz)

Introduction

- ✔ 1)The Kohonen algorithm (SOM)
- ✔ 2) Forecasting vectors
- ✔ 3) Study of trajectories
- ✔ 4) Ozone pollution

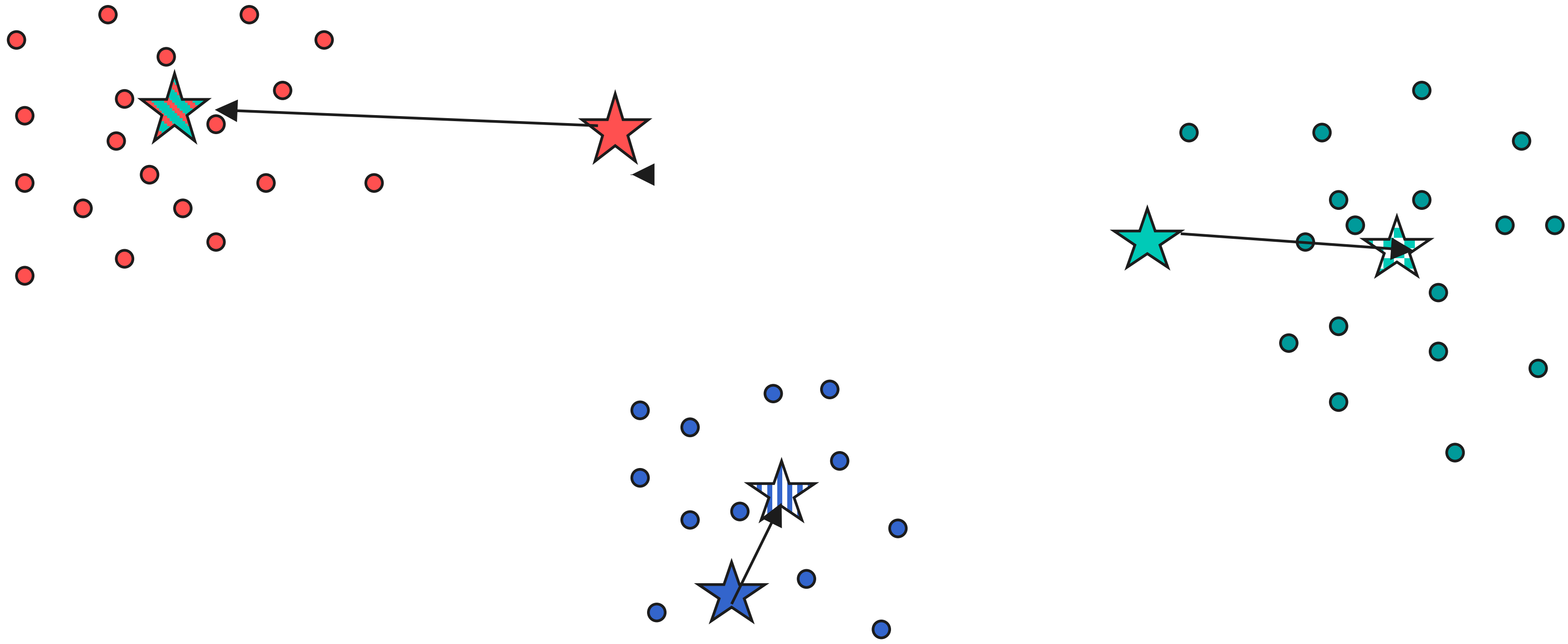
Kohonen algorithm vs classical classification

- The classical classification algorithms are
  - the Forgy algorithm (or moving centers algorithm)
  - the ascending hierarchical algorithm
- (+ variants)
- Both are deterministic
- Two main differences :
  - The SOM algorithm is stochastic
  - A neighborhood structure between classes is defined

Forgy algorithm

At each step, the classes are defined (by the nearest neighbor method)

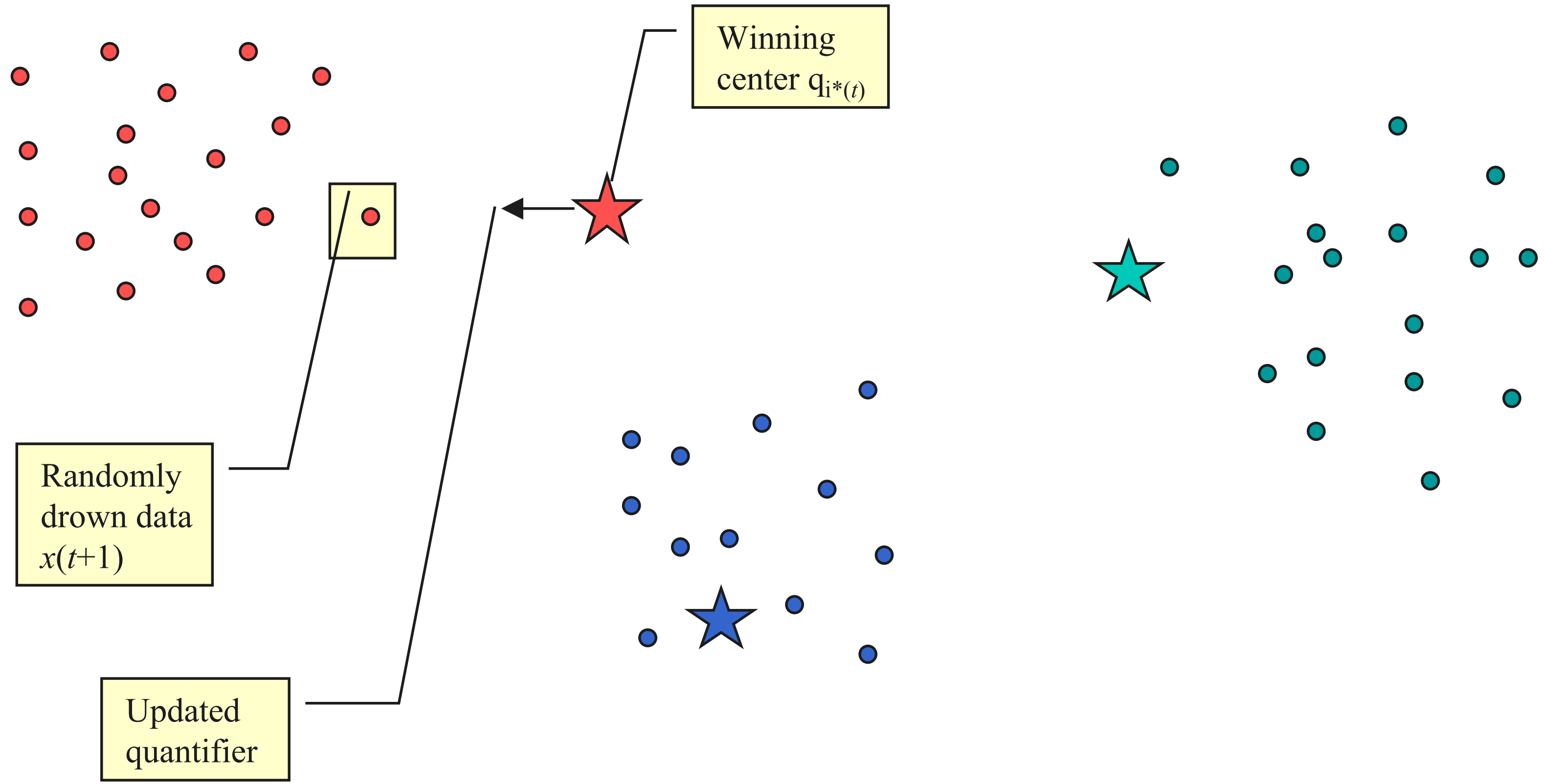
The code vectors are updated to be placed at the gravity center of the classes, etc.



After randomly choosing the code vectors, the associated classes are defined, then the classes are determined, then the code vectors and so on

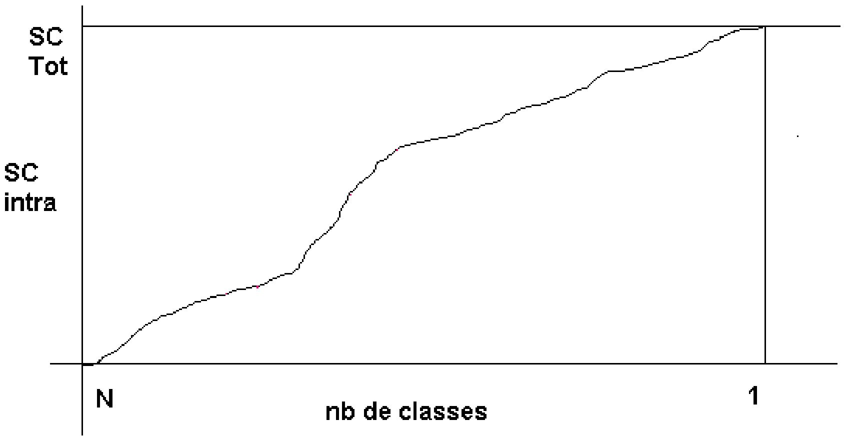
Competitive learning (without neighborhood)

- There exists a stochastic version of the Forgy algorithm, which is exactly the Kohonen algorithm without neighbor

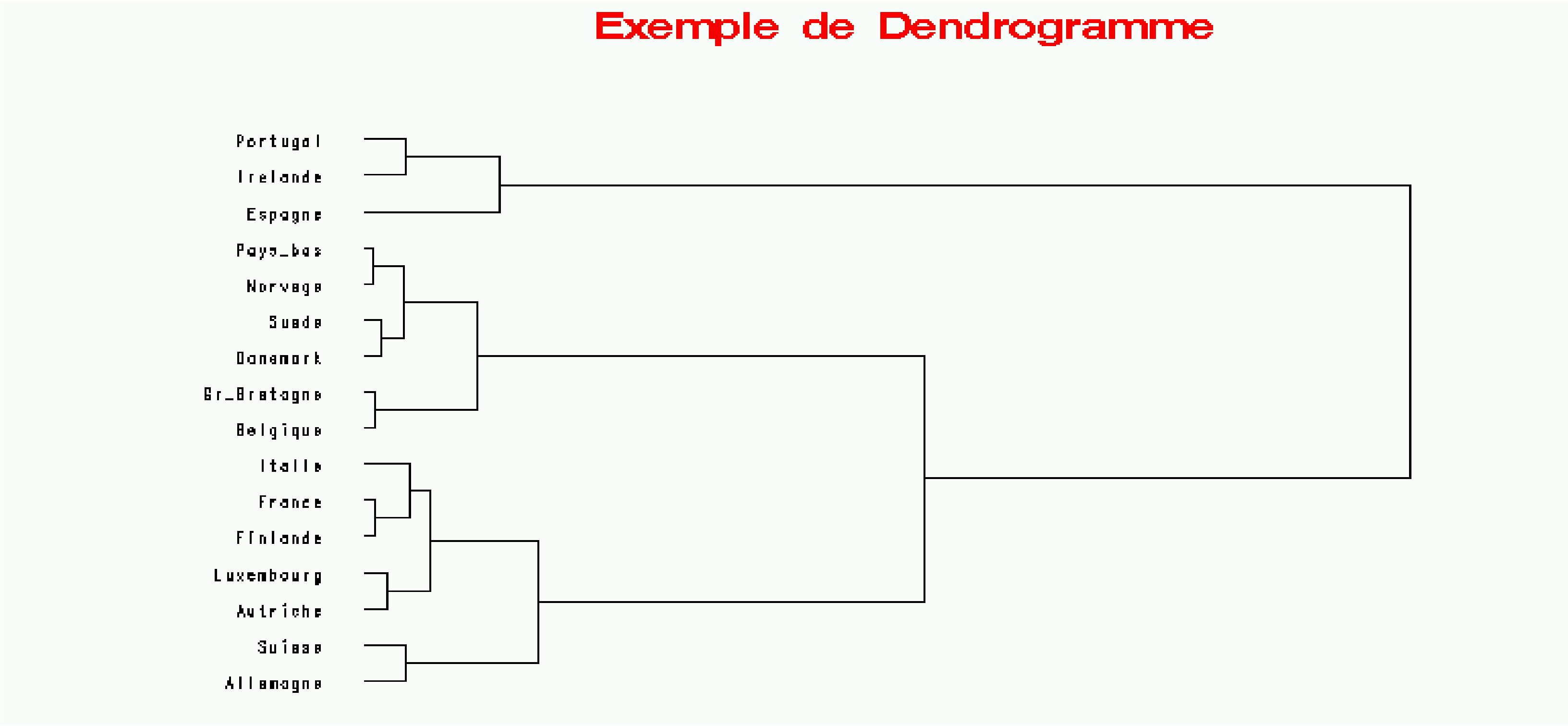


Hierarchical classification

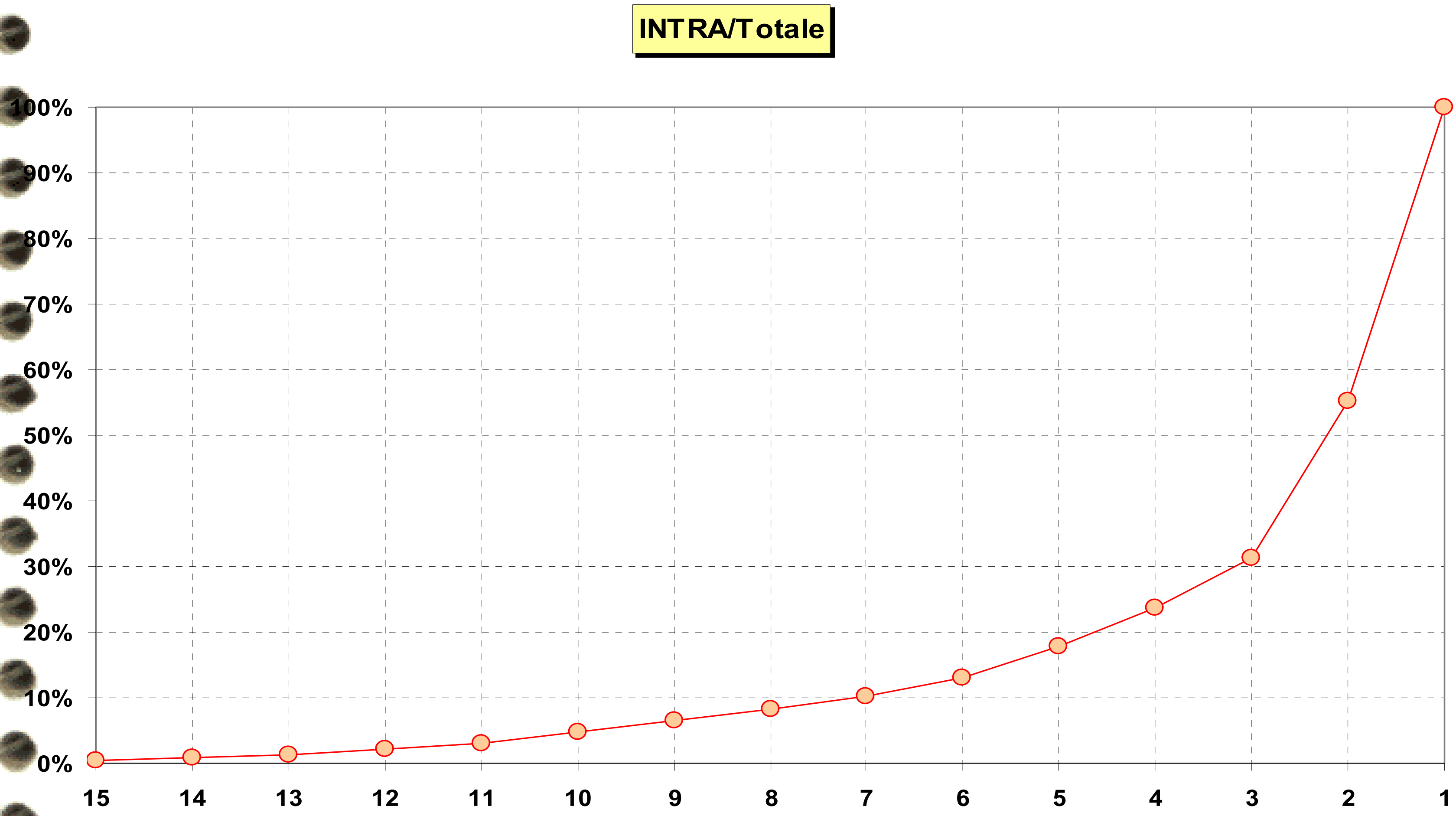
- One builds a sequence of embedded classifications, by grouping the nearest individuals, then the nearest classes, etc. for a given distance
- During the clustering process, the intra-classes sum of squares increases from 0, to the total sum of squares
- In general, one chooses the Ward distance, which minimizes at each step the jumps of the intra-classes sum of squares.



Classification tree



Variation of the intra-classes sum of squares



Number of classes decreasing from 15 to 1

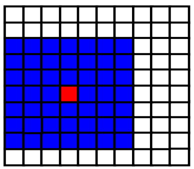
Stochastic vs deterministic

- The Forgy algorithm is the deterministic algorithm associated to the Competitive learning algorithm (algorithm in mean)
- In the same way, the Batch Kohonen algorithm is the mean algorithm associated to the Kohonen algorithm
- The stochastic algorithms have interesting properties,
  - they are on-line algorithm
  - they can escape from some of the local minima

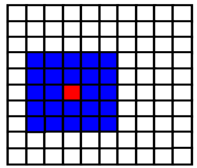
Some neighborhood structures

- One has to define a neighborhood structure among the classes

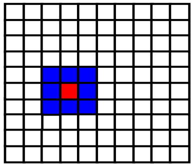
Grid



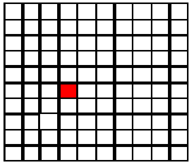
Voisinage de 49



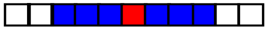
Voisinage de 25



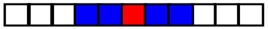
Voisinage de 9



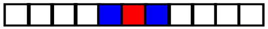
Voisinage de 1



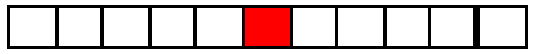
Voisinage de 7



Voisinage de 5

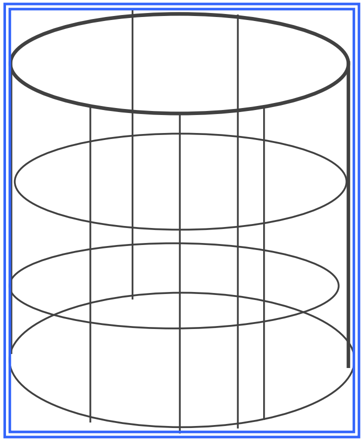


Voisinage de 3



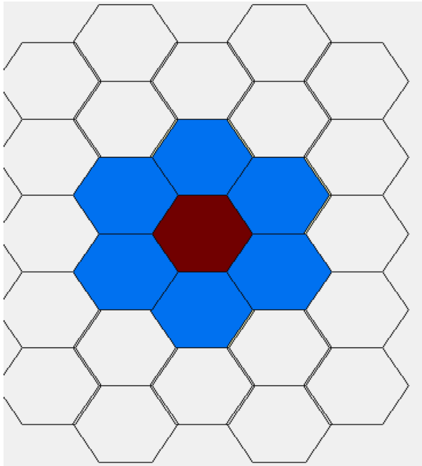
Voisinage de 1

String



Cylinder

Hexagonal



Main property : Self-organization

- If two observations are similar
  - they belong to the same class (property shared by all the classification algorithms)

  OR

  - they belong to neighbor classes
- This organization is not supervised

Mathematical definition

- It is an original classification algorithm, defined by Teuvo Kohonen, in the 80s.
- The algorithm is **iterative**.
- The initialization gives a code-vector to each class, the code-vectors belong to the data space and are randomly chosen

- At each step, an observation is randomly drawn
- It is compared to all the code-vectors
- The **winning class** is defined (its code-vector is the nearest for a given distance)
- **The code-vectors of the winning class and of the neighbor classes are modified in order to be closer to the observation**
- It is an extension of the Competitive Learning algorithm (which does not consider neighborhood)
- It is also a competitive algorithm

Notations

- The data space is K, subset of R^d
- There are n classes, (or n units), structured into a network with predetermined topology (dimension 1, 2, cylinder, torus, hexagonal)
- This structure defines the neighborhood relations, the weight of the neighborhood is defined by a neighborhood function
- The code vector of unit i is denoted C_i, it has d components

- After the random initialization of the code-vectors
- At step t,
  - An observation $x(t+1)$ is drawn
  - The winning unit is denoted $i_0(x(t+1))$
  - The code-vector $C_{i_0(x(t+1))}$ and its neighbors are updated

Definition of the algorithm

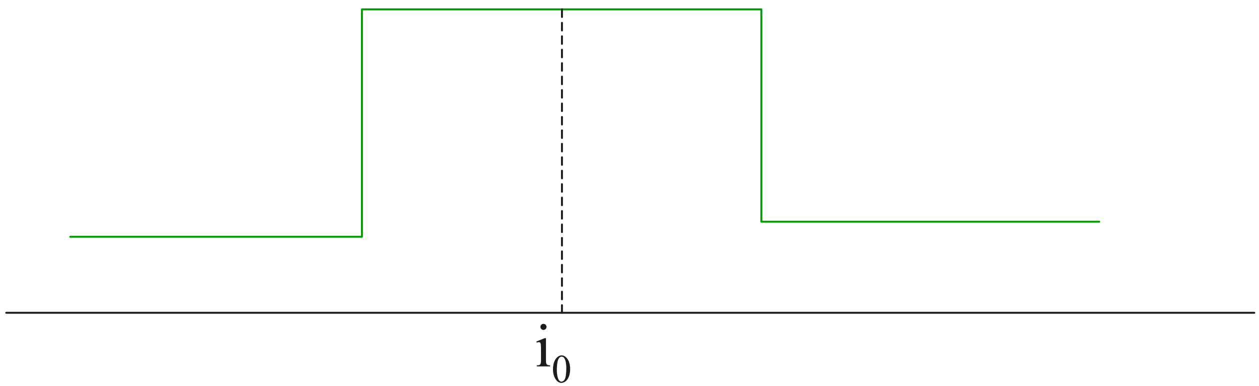
- $\varepsilon(t)$ is the **adaptation parameter**, positive, <1, constant or slowly decreasing
- The **neighborhood function** $\sigma(i,,j)$=1 iff i and j are neighbor, decreasing with $|i\text{-}j|$, the neighborhood size slowly decreases with time
- Two steps, after drawing $x(t+1)$, (independent drawings)
  - Compute the winning unit

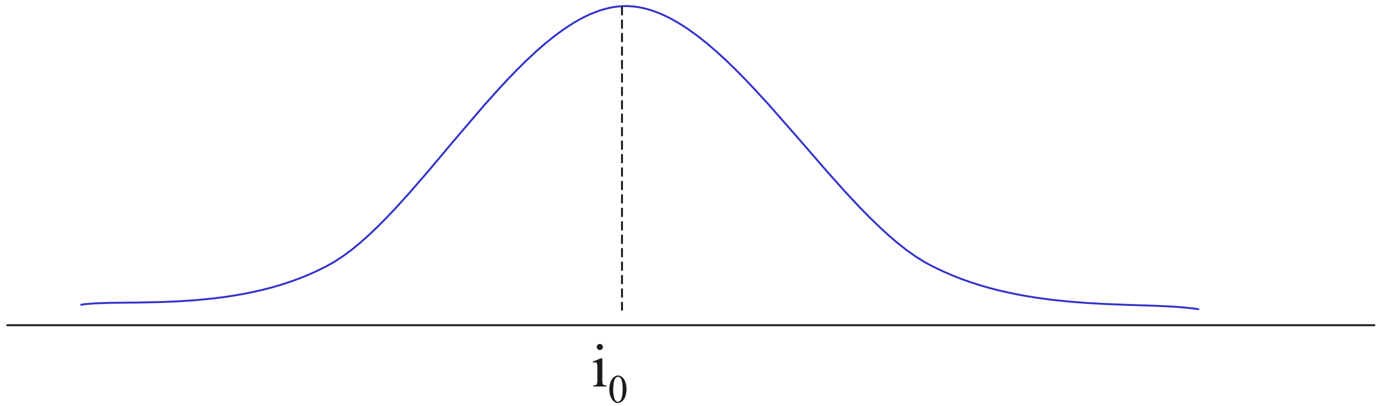
$$i_0(t+1) = \arg\min_i \left\| x(t+1) - C_i(t) \right\|$$

  - Update the code-vectors

$$C_i(t+1) = C_i(t) + \varepsilon(t+1) \quad \sigma(i_0(t+1), \; i)(x(t+1) - C_i(t))$$

Neighborhood functions σ





Theoretical analysis

- The algorithm can be written

$$C(t+1) = C(t) + \varepsilon H(x(t+1), C(t))$$

- The expression looks like a gradient algorithm
- But if the input distribution is continuous, the SOM algorithm is not a gradient algorithm (ERWINN)

- **But in all our applications the data space is finite (data analysis). In this case, there exists an energy function which is an extension of the intra-classes sum of squares** (cf Ritter et al. 92).

- **The algorithm minimizes the sum of the squared distances of each observation not only to its code-vector, but also to the neighbor code-vectors**

Intra-classes sum of squares

- The algorithm SCL (0-neighbor) is the stochastic gradient algorithm which minimizes the intra-classes sum of squares (called quadratic distortion)

$$D(x) = \sum_{i=1}^{n} \sum_{x \in A_i} \left\| x - C_i \right\|^2$$

- A_i is the class represented by the code vector C_i

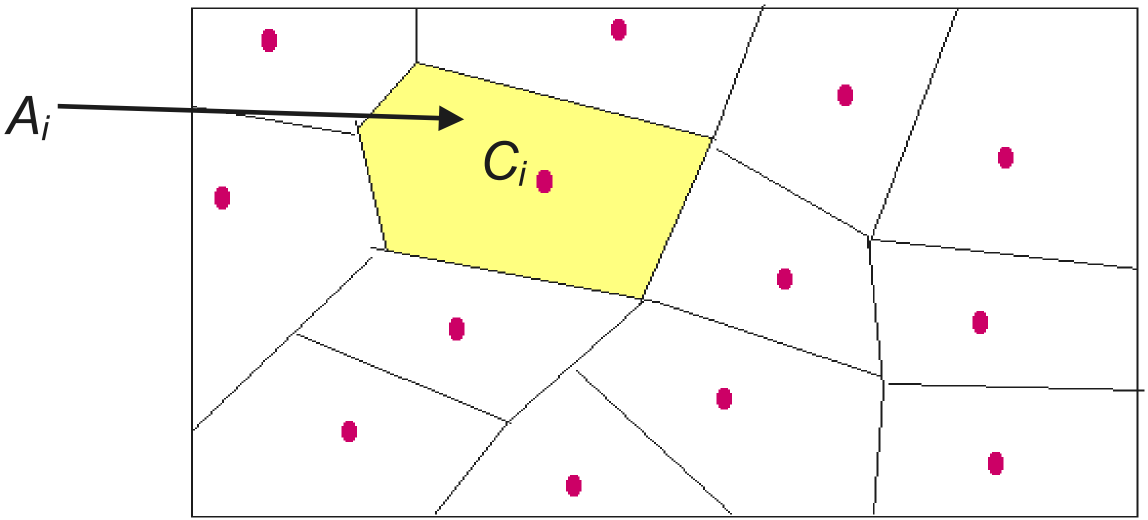
Intra-classes sum of squares extended to the neighbor classes

$$D_{SOM}(x) = \sum_{i=1}^{n} \sum_{\substack{x \text{ s.t.} \\ i = i_0(x) \\ \text{or } i \text{ neighbor of } i_0(x)}} \left\| x - C_i \right\|^2$$

- ✔This function has many local minima
- ✔The algorithm converges, with Robbins-Monro hypothesis on the ε, (they have to decrease neither too slowly, nor too quickly)
- ✔***The complete proof is available only for a restricted case, (dimension 1 for the data, dimension 1 for the structure).***
- ✔To accelerate the convergence, the size of the neighborhood is large at the beginning and decreasing.

Voronoï classes

- *In the data space, the classes provide a partition, or Voronoï mosaic,which depends on the C_i.*
- $A_i(C) = \{x \ / \ ||C_i - x|| = \min_j || C_j - x || \}$: *i*-th class. Its elements are the data for which C_i is the winning code-vector.



C_i is the code-vector of class A_i

What it does ?

- The SOM algorithm groups the observations into classes
- Each class is represented by its code-vector
- Its elements are similar between them, and resemble the elements of neighbor classes
- This property provides a nice visualization along a Kohonen map

Clustering Kohonen classes

- ✔ The number of classes has to be pre-defined, it is generally large
- ✔ So it is very useful to reduce the number of classes, by using a hierarchical clustering. This second clustering groups only contiguous classes (for the organization property)
- ✔ This fact gives interesting visual properties on the maps.

Applications for temporal data

- Many applications of the Kohonen algorithm to represent high dimensional data
- The purpose is to give some examples of applications to temporal data, data for which the time is important
- Rousset, Girard (consumption curves)
- Gaubert (Panel Study of Income Dynamics in USA (5000 households from 1968)
- Rynkiewicz, Letrémy (Pollution)

Forecasting for vectorial data with fixed size

- Problem : predict a curve (or a vector)
- Example : a consumption curve for the next 24 hours, the time unit is the hour and one has to simultaneously forecast the 48 values of the complete following day (data from EDF, or from Polish consumption)
- First idea : to use a recurrence
  - Predict at time *t*, the value X_{t+1} of the next half-hour
  - Consider this predicted value as an input value and repeat that 48 times
- PROBLEM :
  - with ARIMA, crashing of the prediction, which converges to a constant depending on the coefficients
  - with neural non linear model, chaotic behavior due to theoretical reasons
- New method based on Kohonen classification

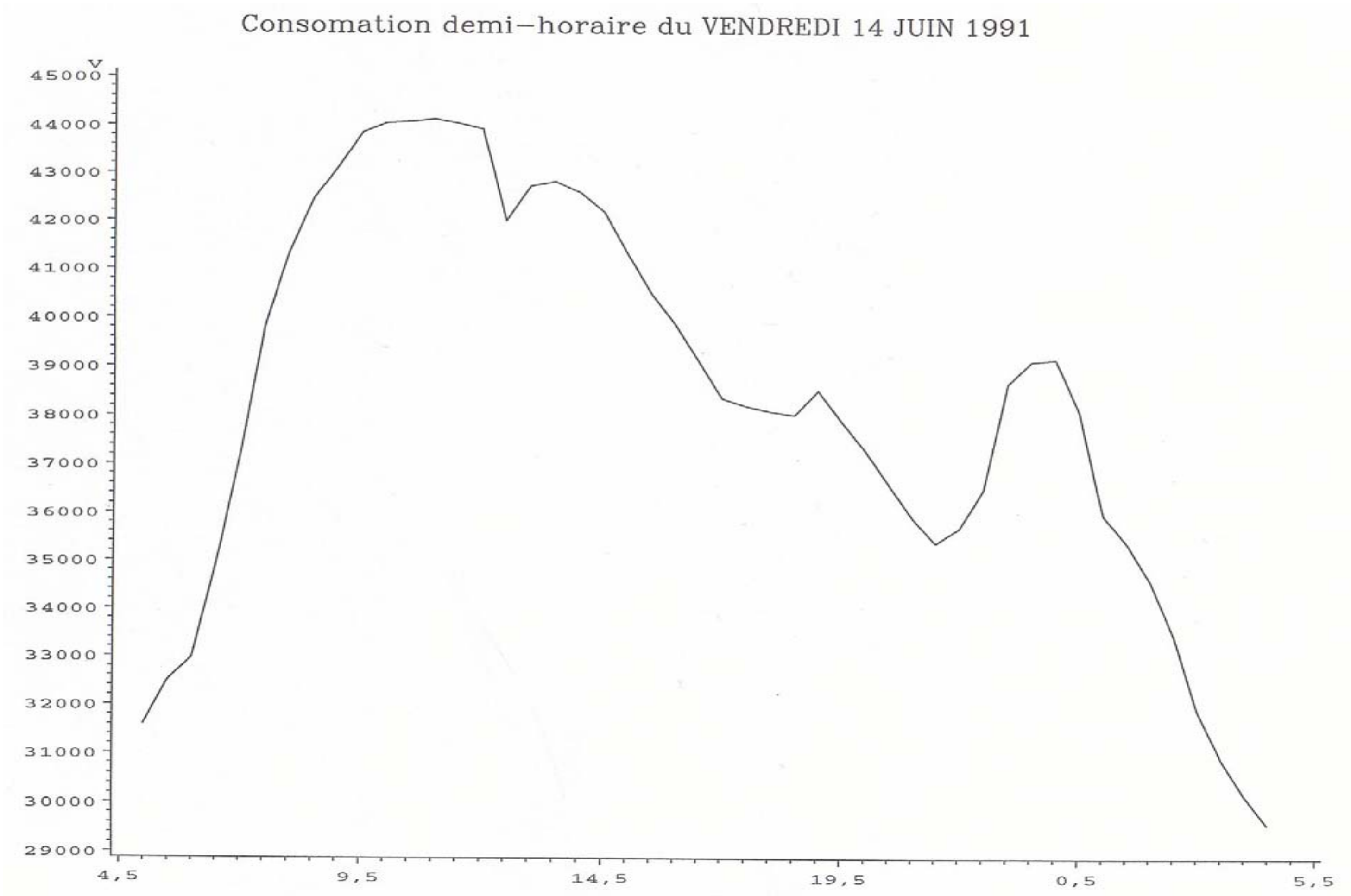
The data

The power curves are quite different from one day to another

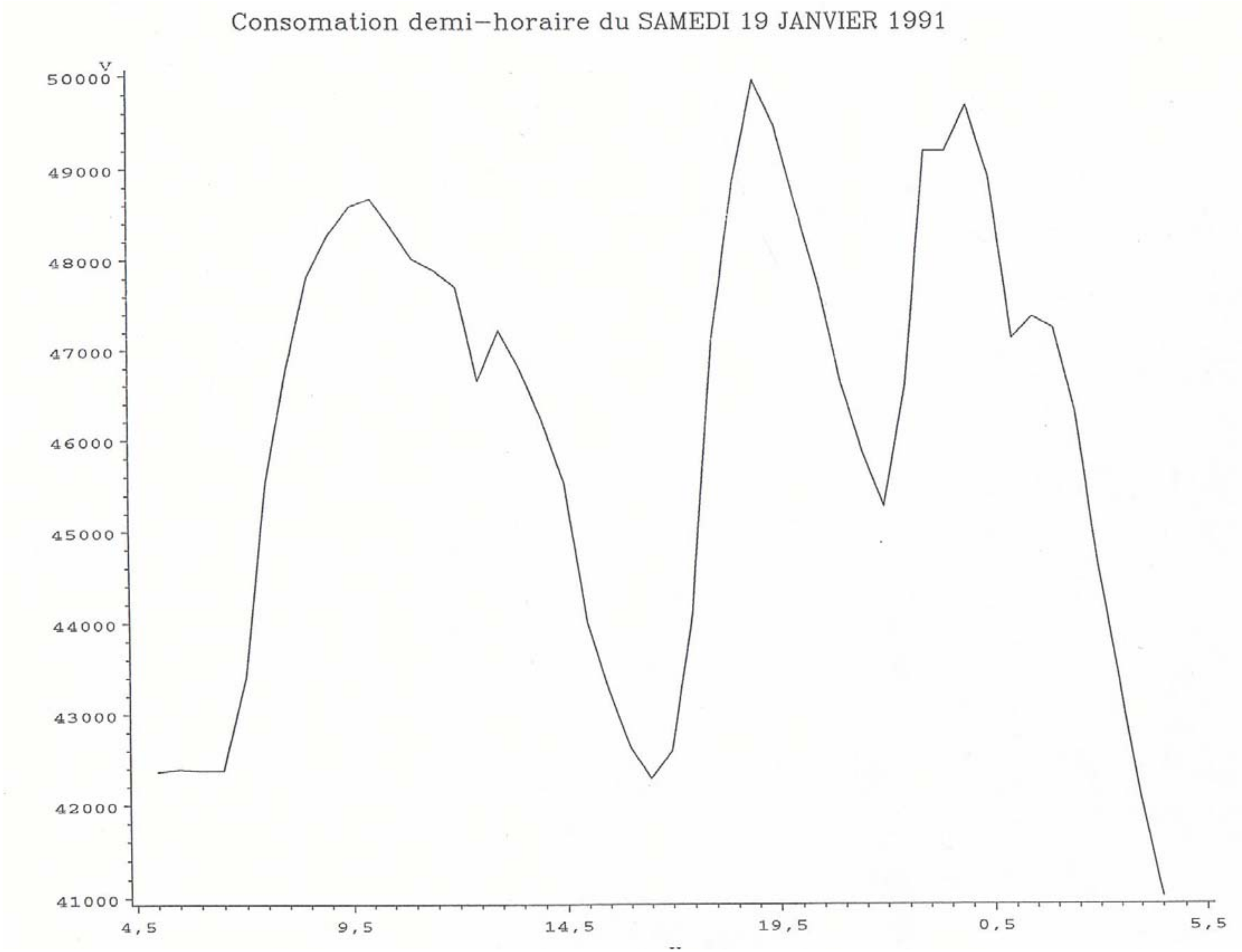
It strongly depend on

- the season
- the day in the week
- the nature of the day (holiday, work day, saturday, sunday, EJP, ...)

Shape of the curves



Shape of the curves



Method

- Decompose the curve into three characteristics

the mean m, the variance σ^2, the profile P defined by

$$P(j) = \left(P\left(j,h\right), h = 1, \cdots, 48\right) = \left(\frac{V\left(j,h\right) - m\left(j\right)}{\sigma\left(j\right)}\right)$$

j is the day, *h* is the half-hour

- Predict the mean and the variance (one dimensional prediction)

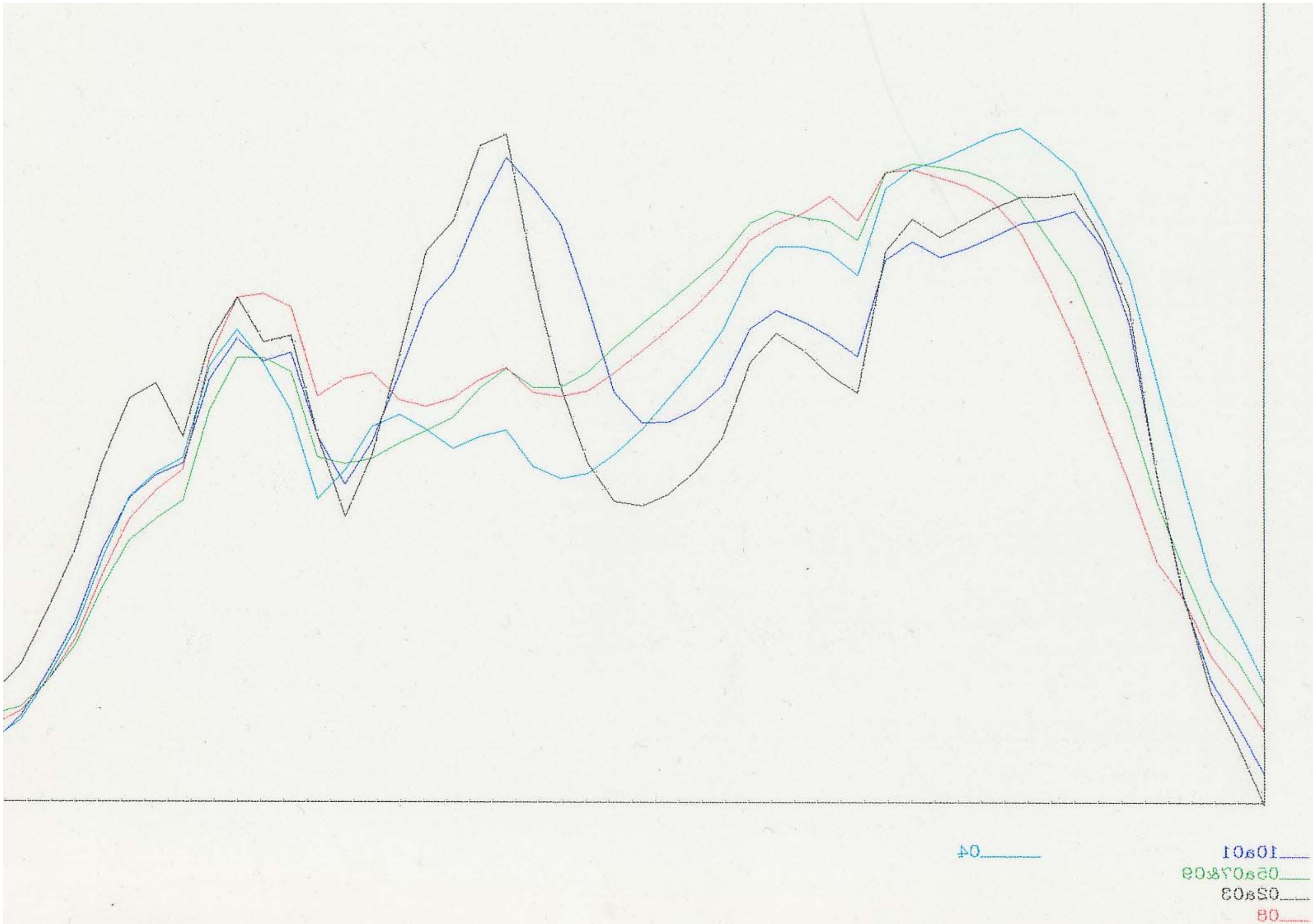
- Achieve a classification of the profiles

- For a given unknown day, build its typical profile and redress it (multiply by the standard deviation and add the mean)

Method

- The mean and the variance are forecast with an ARIMA model or with a Multilayer Perceptron
- The input variables are some lags, meteo variables, nature of the day
- The 48 - vectors are normalized to compute the profile : their norms are equal to 1.
- The origin is taken at 4 h 30 : the value at this point is relatively stable from one day to another

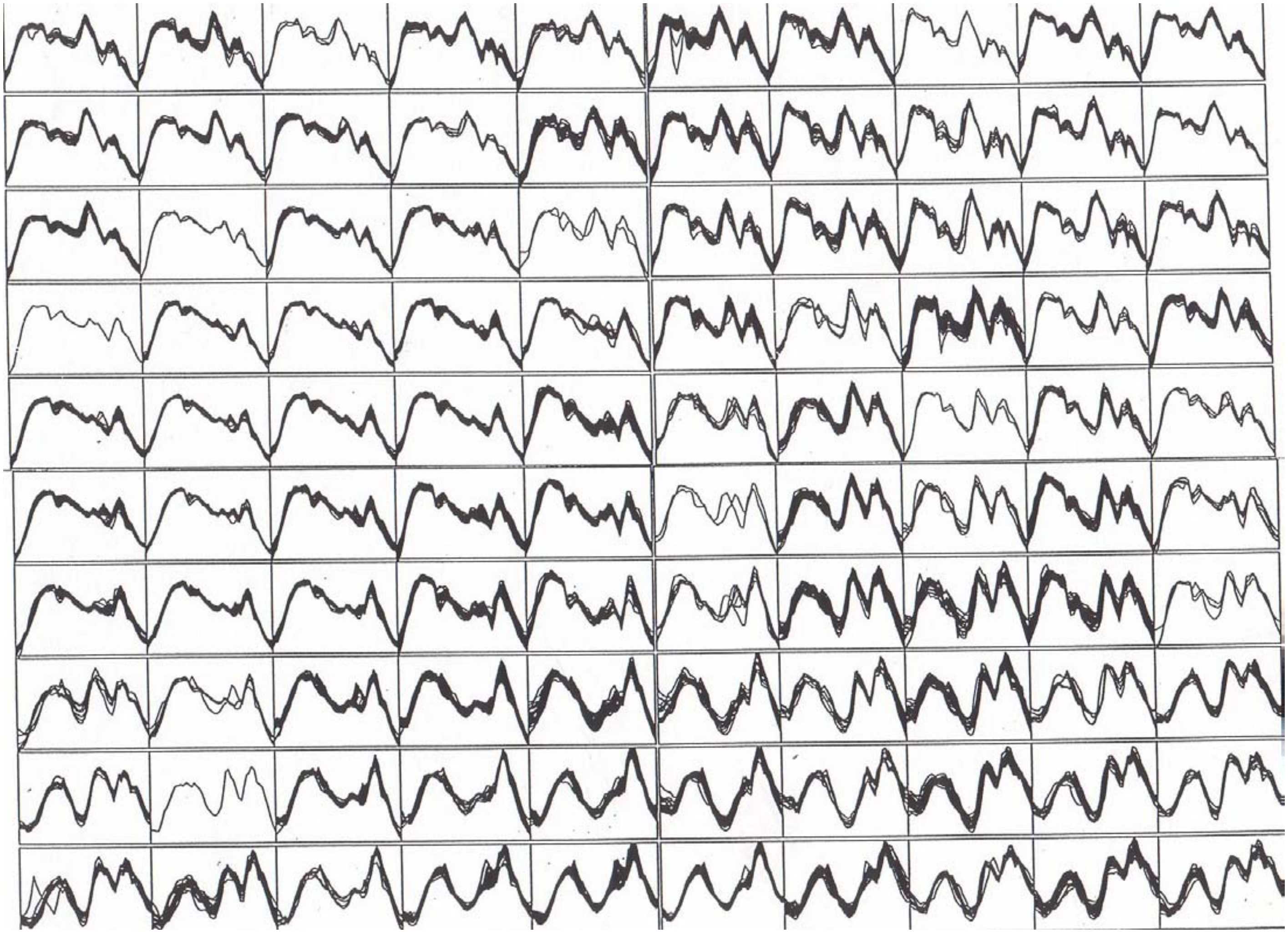
Origin of the day



The profiles

- ✔ The distance between two profiles is computed with the same weight for each half-hour
- ✔ The weather does not influence the profile : it acts only on the mean and the variance
- ✔ Classification of the profiles, (vectors in R^{48}, with norm 1, and sum 0)
- ✔ Classification using the Kohonen algorithm

Classification of the profiles

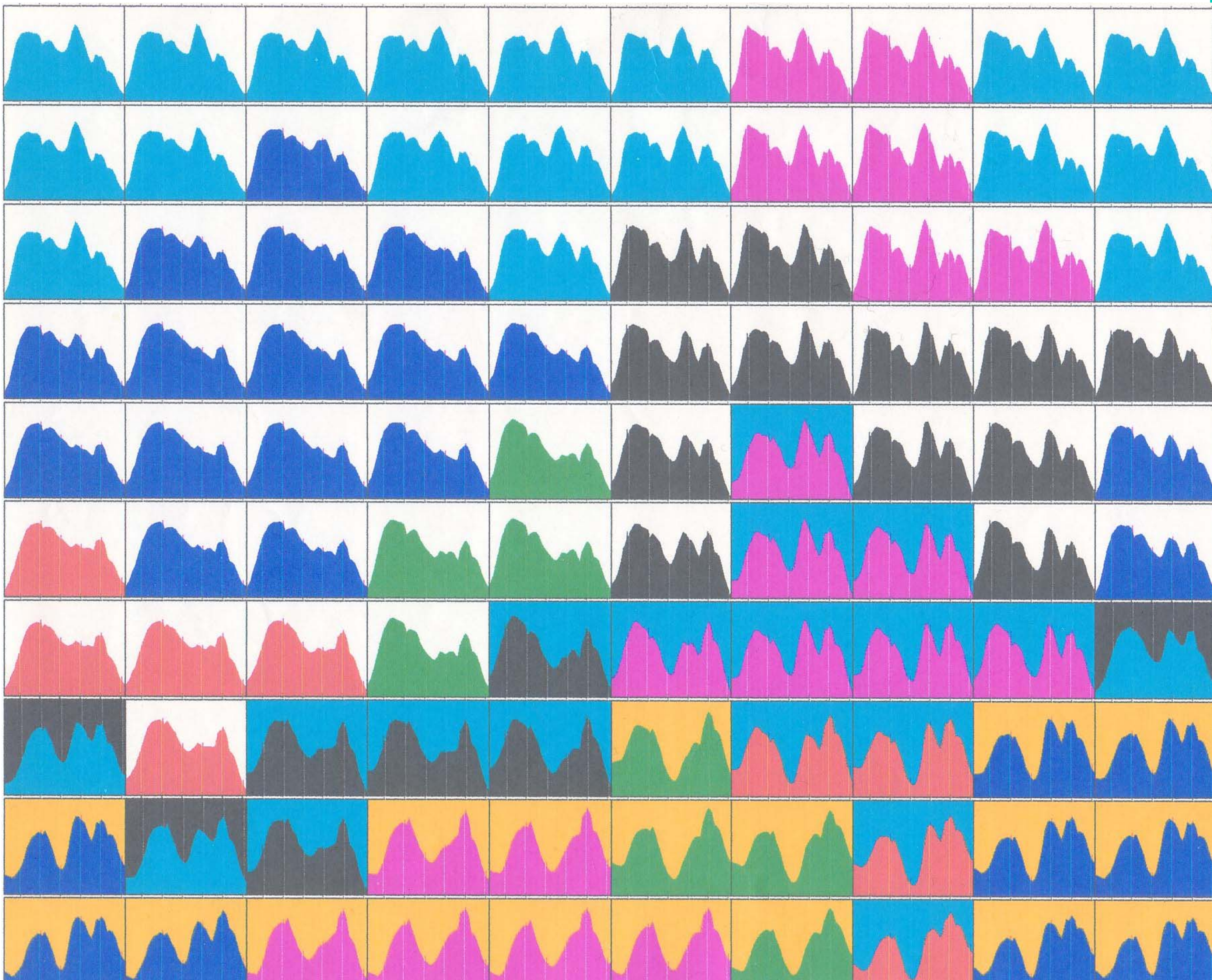


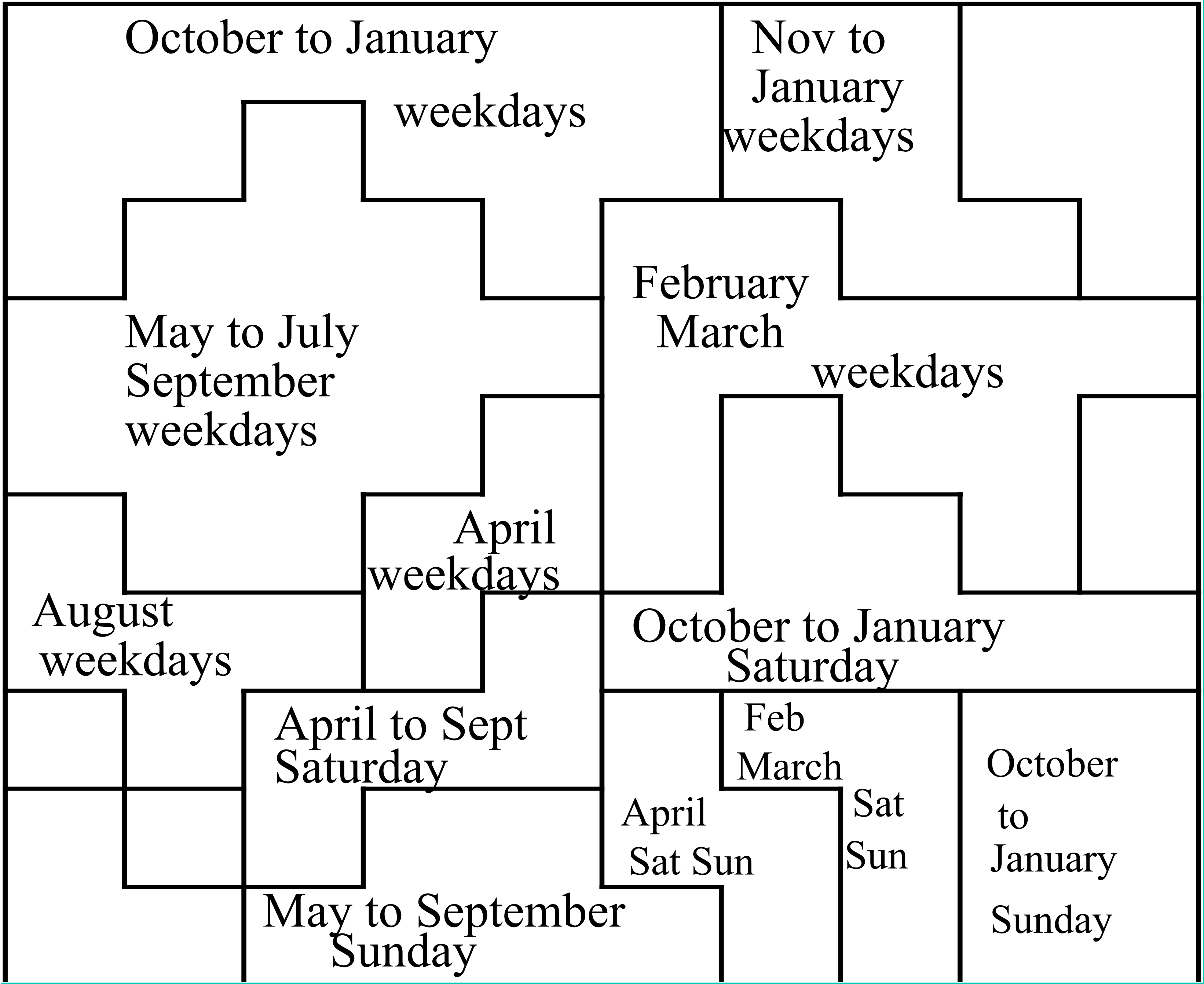
Advantages of the Kohonen method

- Advantages of the Kohonen algorithm
  - The similar vectors belong to neighbor classes
  - The typical profile is chosen as representative of the class
  - It is very simple to go to on the computation on new data, starting from the last values of the weights

Clustering the classes

- To facilitate the interpretation of the classes, the 100 classes are grouped into 13 classes, according to a hierarchical classification
- The limits of the new classes corresponds to the greatest inter-classes distances for the 100-classes classification
- One can observe that there is a significant arrangement on the map : from the top to the bottom, one can encounter successively the weekdays of Autumn and Winter, the weekdays of Spring and Summer, and the Saturdays and Sundays
- These super classes are only used for representation




October to January
weekdays
Nov to January weekdays
May to July September weekdays
February March
weekdays
April weekdays
August weekdays
October to January Saturday
April to Sept Saturday
Feb March
Sat Sun
April Sat Sun
October to January Sunday
May to September Sunday

Using for forecasting

- To use this classification
  - classify the past days as before
  - make a calendar for associating to a given day *j* the number *i*(*j*) of a class (or eventually the numbers of all the classes which contains this day), with their repetitions
  - forecast the mean and the standard deviation with a one dimensional method, (ARIMA or perceptron) for the day *j*
  - the forecasted curve for the day *j* is the profile associated to the class *i*(*j*) , (i.e. the mean profile of this class), or the *weighted mean of the profiles of the concerned classes*, corrected by multiplying by the standard deviation and by adding the mean

Corrected curves

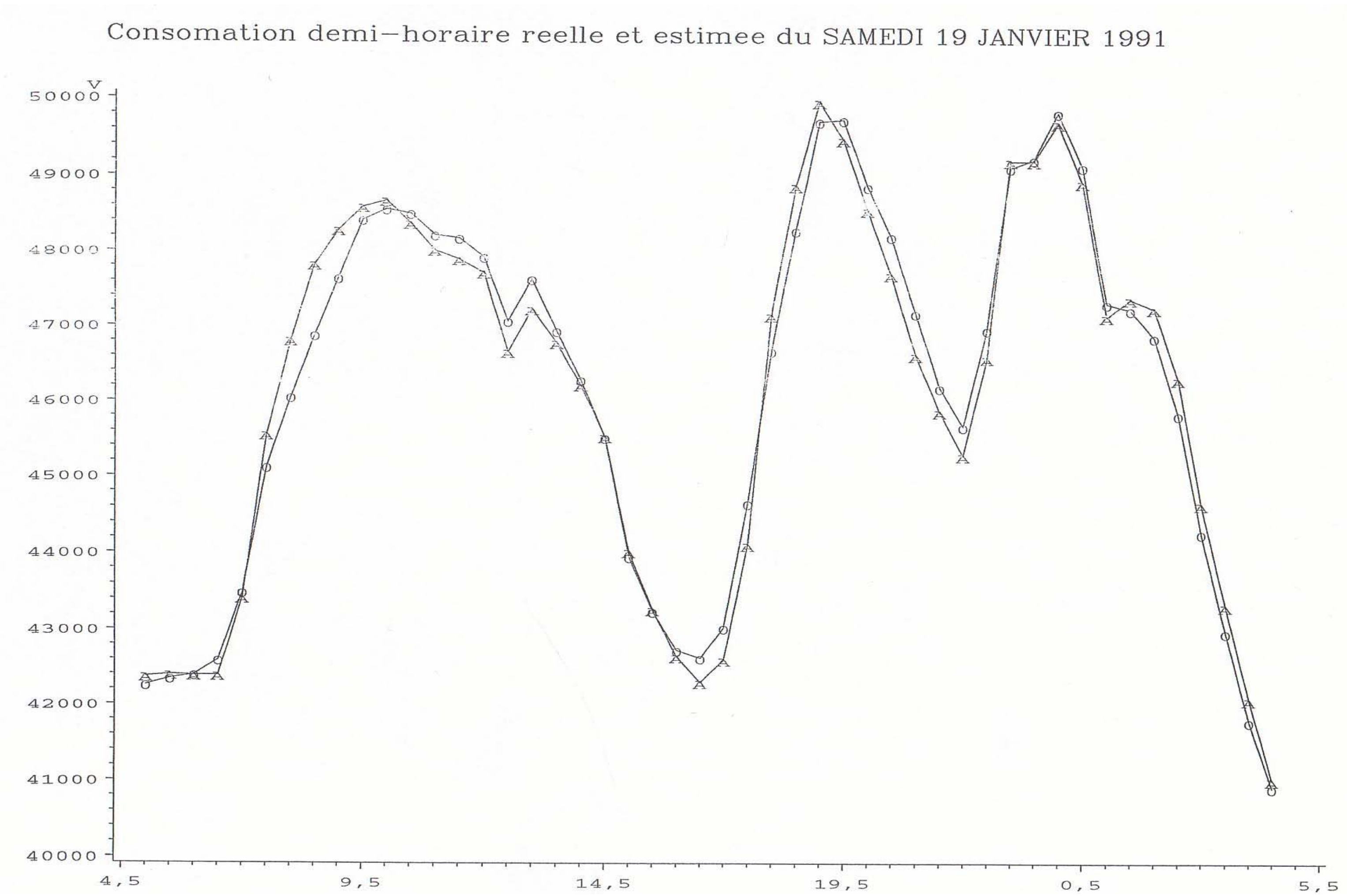
- For a day j, let a_{ji} be the number of instances of the day j in the class i
- Let C_i be the weight vector of the unit i
- The estimated profile of the day j is

$$\hat{P}(j) = \frac{\sum_{i=1}^{100} a_{ji} C_i}{\sum_{i=1}^{100} a_{ji}}$$

- This profile is corrected and the forecasted curve is

$$\hat{V}(j) = \sigma(j)\hat{P}(j) + m(j)$$

Examples of real and forecast curves



Domain of applications

- The classification method is illustrated with the example of the power curves, but it can be used for any classification task
- Electroencephalograms
- Electrocardiograms
- Changes ratio curves
- Control screens
- Price curves
- etc..

- The forecast method is also useful for any kind of curves

Study of individual trajectories

- Let us consider individual data that describe 2507 households by 15 quantitative variables, and for each year from 1984 to 1991.
- So we have (3000 by 8) 15-vectors

- The goal is to produce a robust segmentation using representative variables
- Internal Market / External Market
  - rules governing the relations between the workers and their occupation

- Primary Segment / Secondary Segment
  - Qualitative comparison of existing jobs

- Panel Study of Income Dynamics in USA (5000 households from 1968)

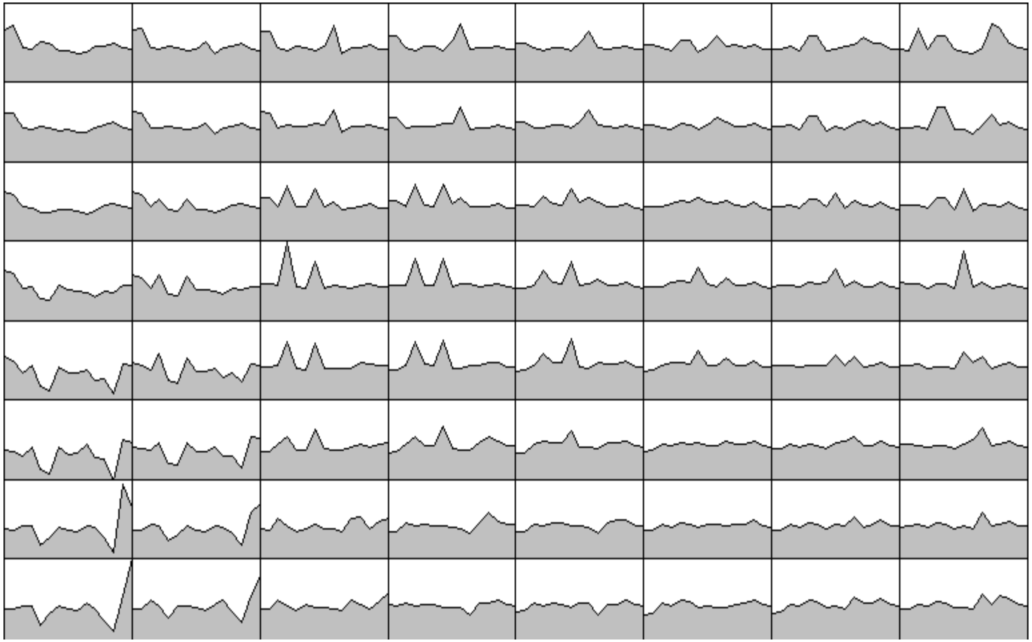
The data (quantitative variables)

- AGEH age of the head of household en 1984.
- ANCH number of years of work since the age of 18.
- CRSALH annual rate of growth of the hourly wage
- HEXJH annual number of work hours in extra jobs.
- HMJH annual number of work hours (main job).
- HWMJH number of hours per week (main job).
- NBXJH number of extra jobs.
- RSALH hourly wage (without the effect of the inflation).
- SENH seniority in the current job.
- TAIFAM size of the family in 1984.
- VHWMJH variation of the number of work hours per week (main job)
- VWMJH variation of the number of work weeks (main job).
- WMJH number of work weeks (main job).
- WOUTH number of weeks out the labor force
- WUNEH number of weeks unemployed (previous year).

Table : The observed or computed quantitative variables

Kohonen Classification

- Kohonen Algorithm , (8, 8) grid
- 2507 heads of households , en 1984, 1988, 1992, without missing values
- Standardized Data Matrix with 15 columns and 7521 rows



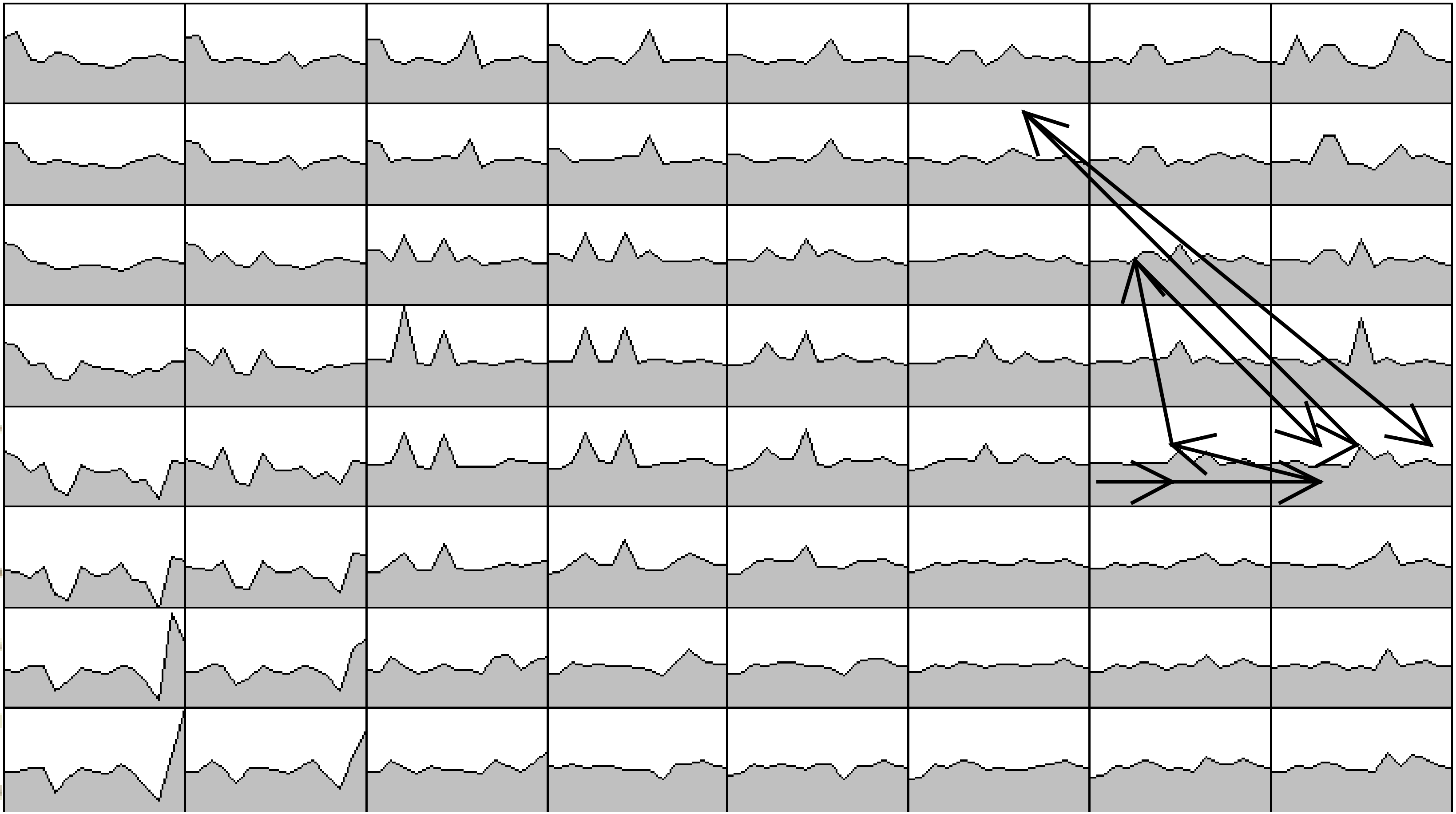
Profiles of the 64 code-vectors

Interpretation of the Grid Classification

- ✔ Main diagonal : quality and and quantity of work increasing from bottom to top)
- ✔ Secondary diagonal : age and seniority (the age decreases from top to bottom), clear opposition between the older workers in the upper left and the younger ones in the lower right.

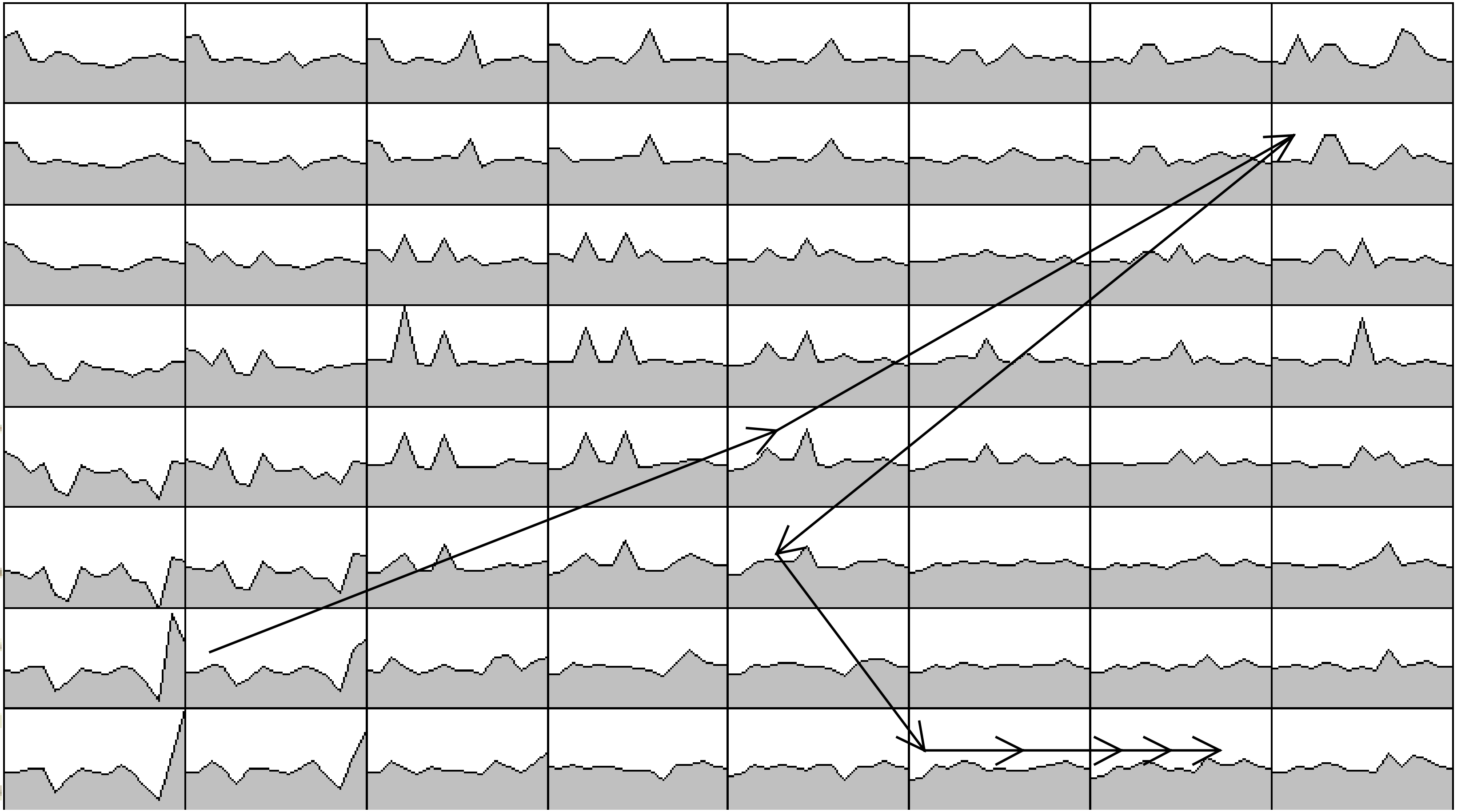
- ✔ In the lower left corner, classes containing individuals with no job (out of the labor force or unemployed) most of the year,
- ✔ In the central region, classes with people exerting more than one job at the same time,
- ✔ In the upper right corner, job situations with stability and high pay.

Trajectories from 1984 to 1992



Individual staying in good job situation during the whole period

Trajectories from 1984 to 1992



Individual leaving the more precarious situation, to reach, after one year in a good situation, an intermediate position

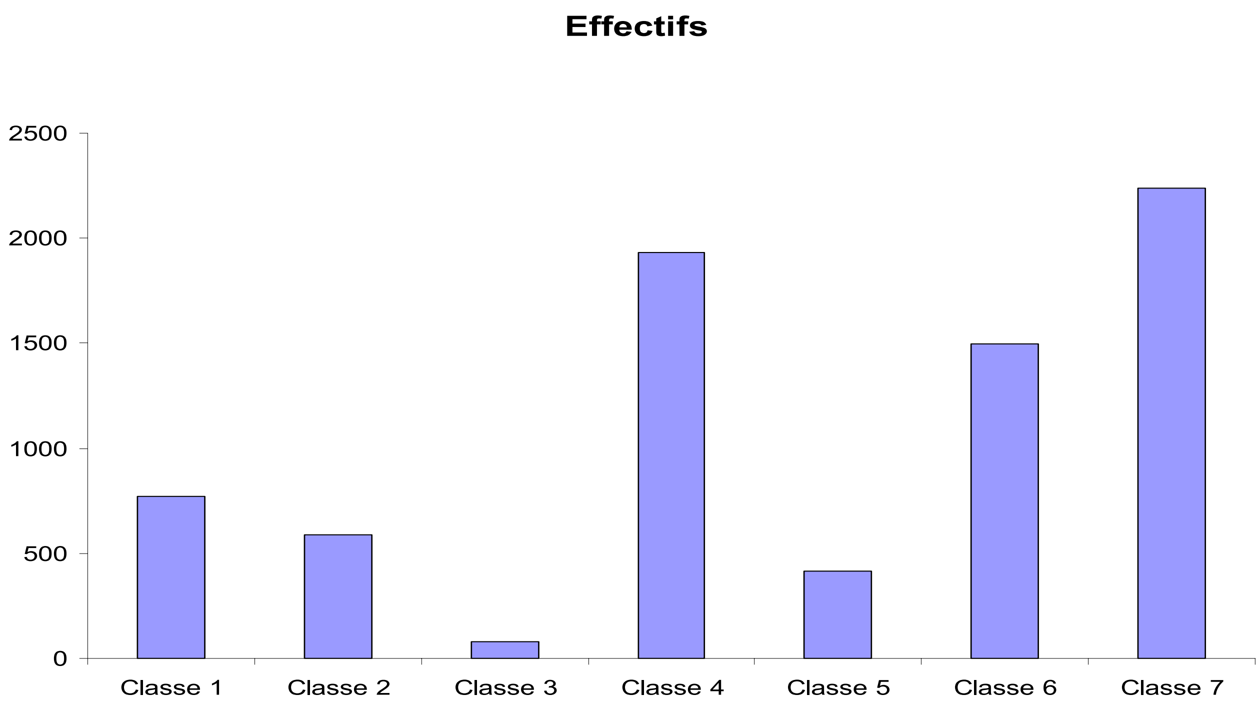
Clustering into 7 classes

Kohonen String on the 64 code-vectors 7 classes

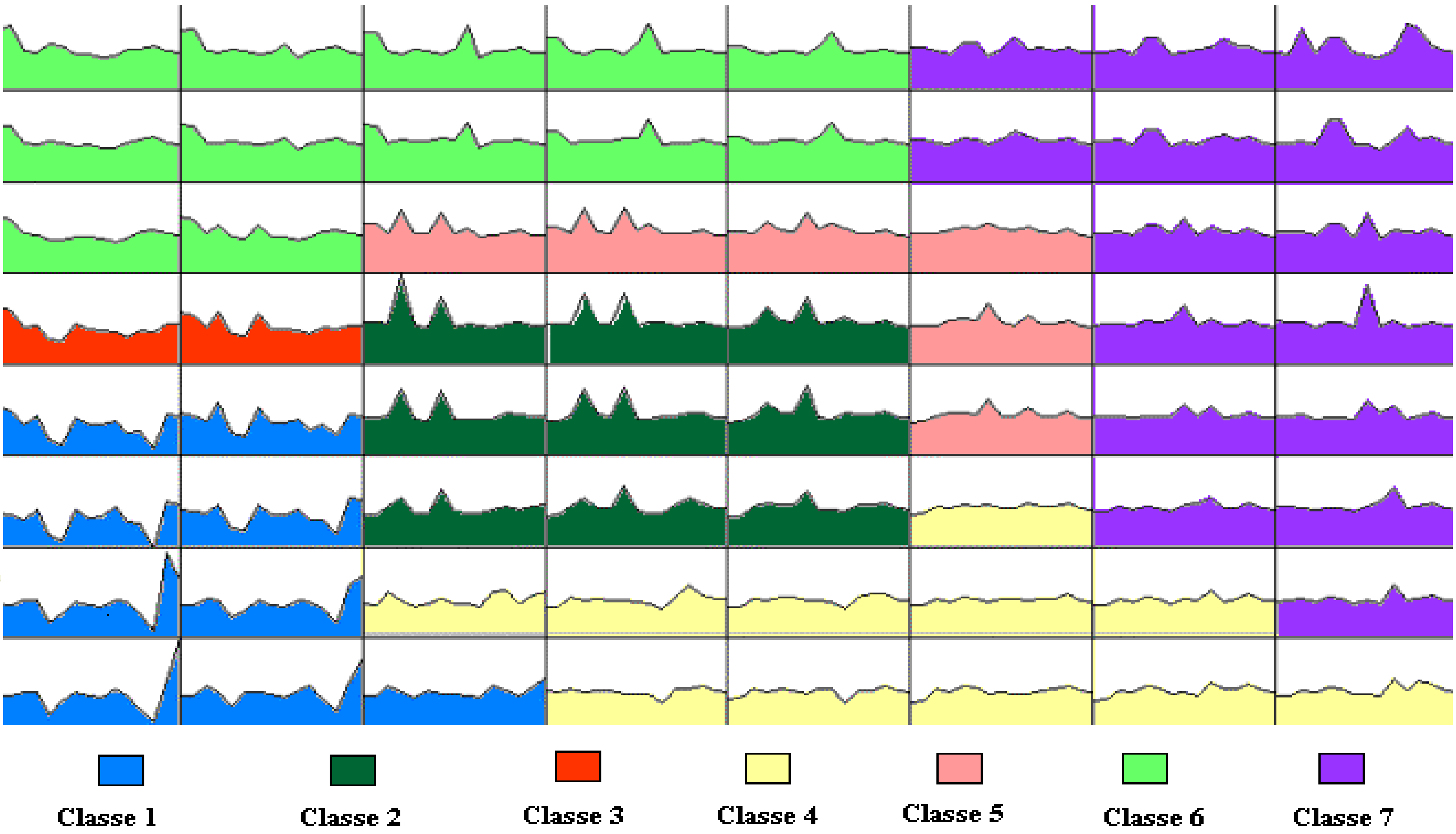
	Population totale	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7
AGEH	40.12	36.4	*35.32*	**59.41**	*33.18*	40.58	**52.69**	39.46
ANCH	15.43	*10.56*	*11.26*	**30.32**	*8.65*	16.18	**28.20**	14.86
CRSALH	0.06	*-0.18*	0.02	0.07	0.06	0.03	0.02	**0.19**
HEXJH	60.70	*12.98*	**562.12**	56.01	*0.25*	**215.01**	*7.39*	*4.74*
HMJH	1974	*663*	1994	*901*	2040	2136	2008	**2348**
HWMJH	42.18	*24.69*	41.88	*22.95*	42.09	44.34	42.09	**48.72**
NBXJH	0.18	*0.05*	**1.24**	0.28	*0*	**1.03**	*0.06*	*0.03*
RSALH	13.35	*6.47*	10.60	10.95	11.30	14.77	13.88	**17.70**
SENH	91.14	*19.51*	64.02	41.05	58.28	118.81	**173.39**	93.04
TAIFAM	3.17	2.92	2.93	*2.04*	2.67	**3.88**	2.57	**4.08**
VHWMJH	0.59	*-6.43*	0.06	*-17*	-0.13	0.23	-0.52	**5.23**
VWMJH	0.65	*-15.66*	**2.77**	-3.83	**3.89**	0.17	1.05	**2.92**
WMJH	44.61	*15.29*	**47.51**	40.81	**48.48**	**48.23**	**47.60**	**48.10**
WOUTH	0.69	**5.76**	*0.09*	1.37	*0.13*	*0.05*	*0.06*	*0.11*
WUNEH	2.09	**16.08**	0.80	3.29	0.40	*0.13*	0.41	0.53
Effectif	7521	772	588	79	1932	416	1495	2240

Table : General Mean and mean by super-class

Absolute frequencies of the 7 classes (Kohonen string)



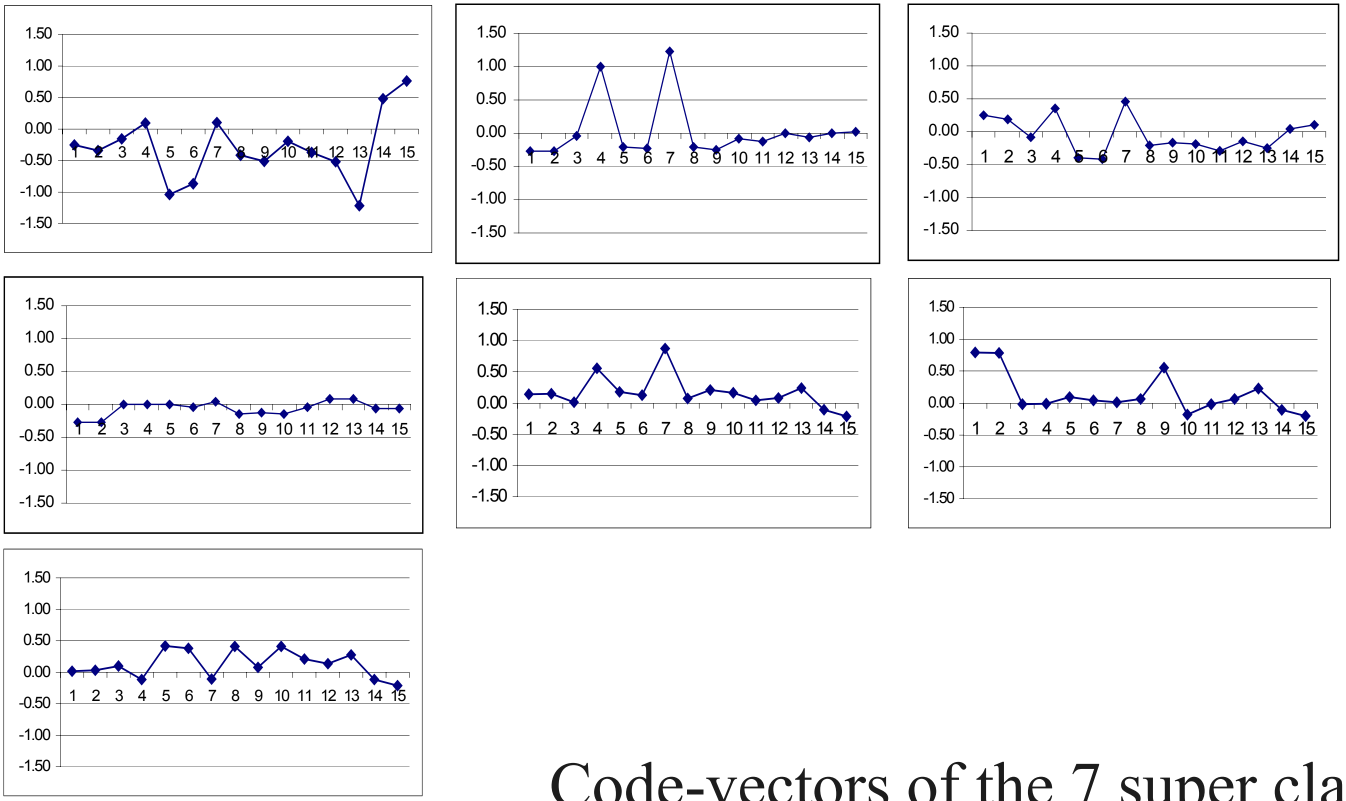
Clustering into 7 classes



Description of the 7 classes

- Class 1 : young, short seniority, less hours, no extra job, low paid, often out of the labor force, negative evolution.
- Class 2 : younger than the average, main full-time job, earnings severely lower than the average, one or more extra jobs.
- Class 3 : old, long seniority, half-time job, low paid, very few extra jobs (close to retirement).
- Class 4 : young, short seniority, no extra job, wages below the average, important augmentation of the number of hours worked.
- Class 5 : one or more extra jobs, with good wages.
- Class 6 : elder, stables, one full time job, earnings close to average.
- Class 7 : middle age, large family (4 persons, one more than average), stables, working a longer duration than the average, without extra job, hourly wages above the average. They have the best growth of their wages and of the work duration.

Description of the 7 classes



Code-vectors of the 7 super classes

Transitions between the 7 classes

Position majoritaire	Effectif	Proba de 1	Proba de 2	Proba de 3	Proba de 4	Proba de 5	Proba de 6	Proba de 7
1	157	**0.75**	0.03	0.01	0.11	0.00	0.03	0.08
2	115	0.04	**0.70**	0.00	0.13	0.07	0.00	0.05
3	10	0.16	0.01	**0.64**	0.01	0.00	0.16	0.02
4	599	0.07	0.06	0.00	**0.77**	0.02	0.00	0.08
5	65	0.01	0.11	0.00	0.01	**0.70**	0.05	0.12
6	498	0.03	0.01	0.02	0.01	0.02	**0.86**	0.06
7	732	0.03	0.02	0.00	0.07	0.03	0.03	**0.82**

Probabilities to be one year in a class, being most of the time in a given class

Pas de position majoritaire	Effectif	Proba de 1	Proba de 2	Proba de 3	Proba de 4	Proba de 5	Proba de 6	Proba de 7
	331	0.14	0.16	0.03	0.22	0.13	0.08	0.23

The same probabilities, when no class has a dominant position

Main result

- The individuals stay most of the time in the same class
- That means that the structure that appears constituted by segments with very different properties with respect to stability, existence of a career, seems to present the quality of a permanent state over a long period
- This will be even clearer with the construction of a Markov chain.
- Except for this latter small group, the less stable classes, relatively, are those corresponding to lower situations, and precisely the two classes having extra job(s).

Clustering into 4 levels

- Classes 1 and 3 are grouped into class A. It is made of the more precarious conditions, recurring unemployment, low pay. Class 3 is not contiguous to class 1 on the string, but it is on the grid. It includes only 25 individuals, so it is reasonable to add it to class 1.

- Classes 2, 4, 5 represent intermediate conditions : important duration of work and moderate wages, 2 or 3 jobs for some of them. They constitute the main class B.

- Classes 6 and 7 are still separated and renamed C and D.

Principal Component Analysis

- PCA on the 15 variables
- 5 axes to get 2/3 of the explained variance
  - (22%, 14%, 11%, 9%, 8%)
- First axis : defined by the variables of activity: the number of work hours, the number of weeks, opposed to the number of weeks of unemployment and out of the labor force.
- Second one opposes age, seniority to the family size (younger family are larger).
- Third one is only defined with the extra job variables.
- The level and the growth of wage and the variables in variation appears only as fourth and fifth axes. That means that the separation of the different situations is mainly explained by other factors than the differentiation of wages.
- Even with this new grouping the main classes are well defined using the 15 quantitative variables . The major characteristics observed above with the more detailed partition are still visible: work duration, seniority, level and growth of real wages, the practice of extra jobs.

Description of the 4 classes

	Population totale	Classe A	Classe B	Classe C	Classe D
AGEH	40.12	38.56	*34.66*	**52.69**	39.46
ANCH	15.43	12.39	*10.24*	**28.20**	14.86
CRSALH	0.06	*-0.15*	0.05	0.02	**0.19**
HEXJH	60.70	16.97	**143.25**	*7.39*	*4.74*
HMJH	1974	*685.26*	2045.05	2008.23	**2348.84**
HWMJH	42.18	*24.51*	42.37	42.09	**48.72**
NBXJH	0.18	0.07	**0.40**	0.06	0.03
RSALH	13.35	*6.88*	11.66	13.88	**17.70**
SENH	91.14	*21.51*	67.99	**173.39**	93.04
TAIFAM	3.17	2.85	2.89	2.57	**4.08**
VHWMJH	0.59	*-7.41*	-0.04	-0.52	**5.23**
VWMJH	0.65	*-14.57*	**3.14**	1.06	**2.92**
WMJH	44.61	*17.66*	48.25	47.60	48.10
WOUTH	0.69	**5.35**	0.11	*0.06*	0.11
WUNEH	2.09	**14.89**	0.44	0.41	0.53
Effectif	7521	851	2936	1495	2240

Mean of the whole sample and by main class A, B, C, D

Frequencies of the qualitative variables

in 1992	Whole sample	Class A	Class B	Class C	Class D
RACE					
1 Whites	69.1 %	51.6	69.2	69.5	74.5
2 Blacks	29.7	48.0	29.8	28.9	23.8
EDUCATION					
1 Primary	0.9	1.4	0.2	2.6	0.4
2 Secondary	18.9	32.7	13.1	27.8	15.1
3 Sec. achieved	40.2	42.7	44.4	37.0	36.9
4 Post-sec.	28.6	18.5	32.7	20.7	32.6
5 BA & more	11.3	4.6	9.5	12.0	15.0
OCCUPATION					
0 No	2.0	17.4	0	0	0.1
1-2 Managers, professionals	36.2	15.7	46.7	32.9	44.9
4 Clerks	12.0	14.2	13.2	14.0	8.8
5 Craftsmen	17.1	14.9	17.0	18.5	17.1
6 Operatives	15.1	14.9	15.1	15.6	14.9
7 Others	12.6	20.6	12.2	15.7	8.5

Example of distribution of some qualitative variables

Transitions between the 4 classes

Position majoritaire	Effectif	Proba de se trouver dans la classe A	Proba de se trouver dans la classe B	Proba de se trouver dans la classe C	Proba de se trouver dans la classe D
A	179	**0.75**	0.13	0.06	0.07
B	951	0.07	**0.82**	0.01	0.10
C	498	0.05	0.04	**0.86**	0.06
D	732	0.04	0.11	0.03	**0.82**

Probabilities to be one year in a class, being most of the time in a given class

Pas de position majoritaire	Effectif	Proba de se trouver dans la classe A	Proba de se trouver dans la classe B	Proba de se trouver dans la classe C	Proba de se trouver dans la classe D
	147	0.34	0.33	0.13	0.29

Probabilities when no class has a dominant position

Transitions

- Over the 2 507 individuals, only 1 028 different trajectories are found, to be compared to the 4^9 possible trajectories, it is clear that a trajectory cannot be conceived as a random process between the four classes.
- Good stability of the situations, the more stable is class C.
- Only transitions A - B, B - D, D - B occur with a significant probability.
- Individuals who do not remain in any class for a long time spend about the third of the time in each of the classes A, B, D,
- and belong only exceptionally to class C.

Transitions between the 4 classes

	AB	AC	AD	**BA**	BC	**BD**	CA	CB	CD	DA	**DB**	DC
Eff	**554**	177	242	**492**	159	**1036**	175	150	262	241	**871**	306
%	**0.12**	0.04	0.05	**0.11**	0.03	**0.22**	0.04	0.03	0.06	0.05	**0.19**	0.07

The frequencies of the transitions

The transitions occur mainly between classes A, B, D.

The number of improvements (transitions AB, BD) is close to the number of deterioration (BA, DB),

The moves from and to Class C are very few.

Class C is separated, is not a step towards the best state, Class D.

Class C could be a more traditional segment.

Precarious jobs (or no job) as in Class A do not lead to the upper segment D.

Possibilities of rotations (in both directions) between the intermediate and upper segments B and D, but without pass through segment C.

Markov Model

- The empirical probabilities (to stay in the same class or to move from one class to another) may be used to build a Markov transition matrix.
- Let M be this matrix.
- This need some important hypotheses concerning the factors influencing the transitions, precisely that the factors which are taken into account are stable over a long period.
- So we can compute the stationary distribution over the 4 classes, (solution of X=XM) and compare it to the observed distributions over the whole period.

	class A	class B	class C	class D
stationary	.106	.363	.209	.322
1984	.138	.400	.181	.281
1988	.110	.381	.199	.309
1992	.112	.356	.203	.329

Markov Matrix

	A	B	C	D
A	0.57	0.24	0.08	0.11
B	0.06	0.78	0.02	0.14
C	0.04	0.14	0.85	0.06
D	0.04	0.04	0.05	0.77

Conclusions

- The observed distributions (for all the years) are very close to the theoretical distribution, as computed with the Markov model
- They become closer along the time
- We get the same conclusions with the (7, 7) transition matrix
- The next thing to study is a more precise examination of the duration in each state, the influence of the qualitative variables, in particular the sector to which belong the jobs for Class C or D, an exact definition of Class C...
- The method allows to build simulated trajectories, to define segments of the whole population

Ozone pollution (in the region Ile-de-France)

- ✔ The time series is the maximum level of pollution due to the presence of ozone in the air, recorded from 1994 to 1997 in the region near Paris

- ✔ The best model seems to be a two-states Hidden Markov Model

- ✔ How to interpret these two regimes ?

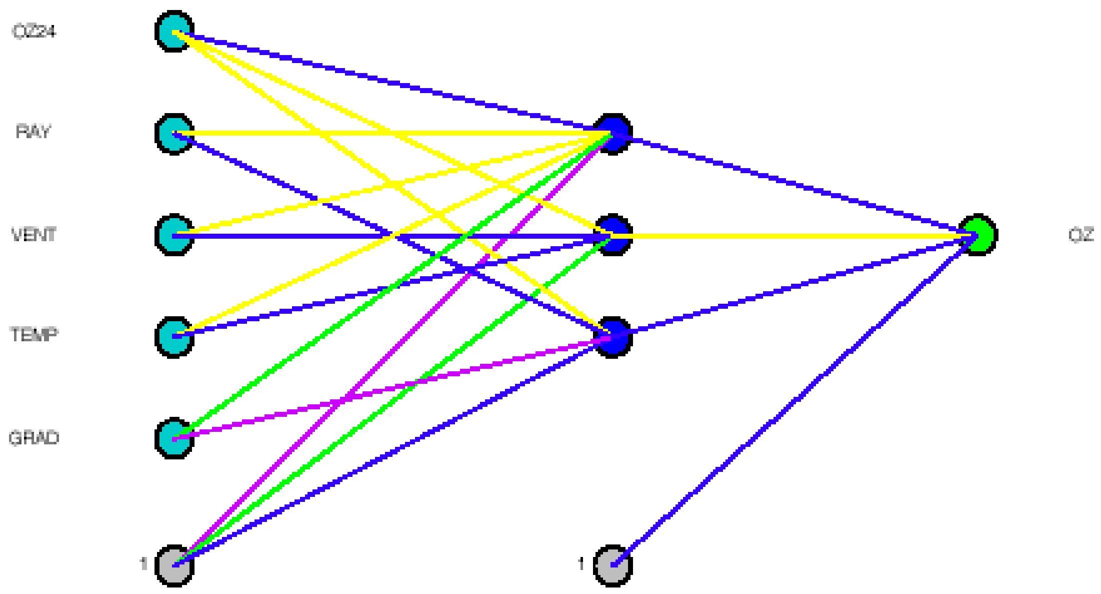
The variables

- ✔ - the maximum of the pollution rate on the day before,
- ✔ - the global radiation,
- ✔ - the mean speed of the wind,
- ✔ - the maximal temperature
- ✔ - the temperature gradient of the day.

- ✔ Two states for the hidden Markov chain, two different auto-regressive models
  - one is linear and is associated to the low or medium values,
  - the second is a Multilayer Perceptron, specialized in the high values.

- ✔ To better understand the nature of both hidden states, the authors classify all the observations (that are 5-dimension vectors) in a 7 by 7 Kohonen map. These 49 classes are grouped into 5 super classes, easy to interpret.

The non linear model (for the high values)



The HMM model

It is possible to estimate the parameters of both models

$$\hat{A} = \begin{pmatrix} 0.97 & 0.02 \\ 0.03 & 0.98 \end{pmatrix}$$

Transition matrix

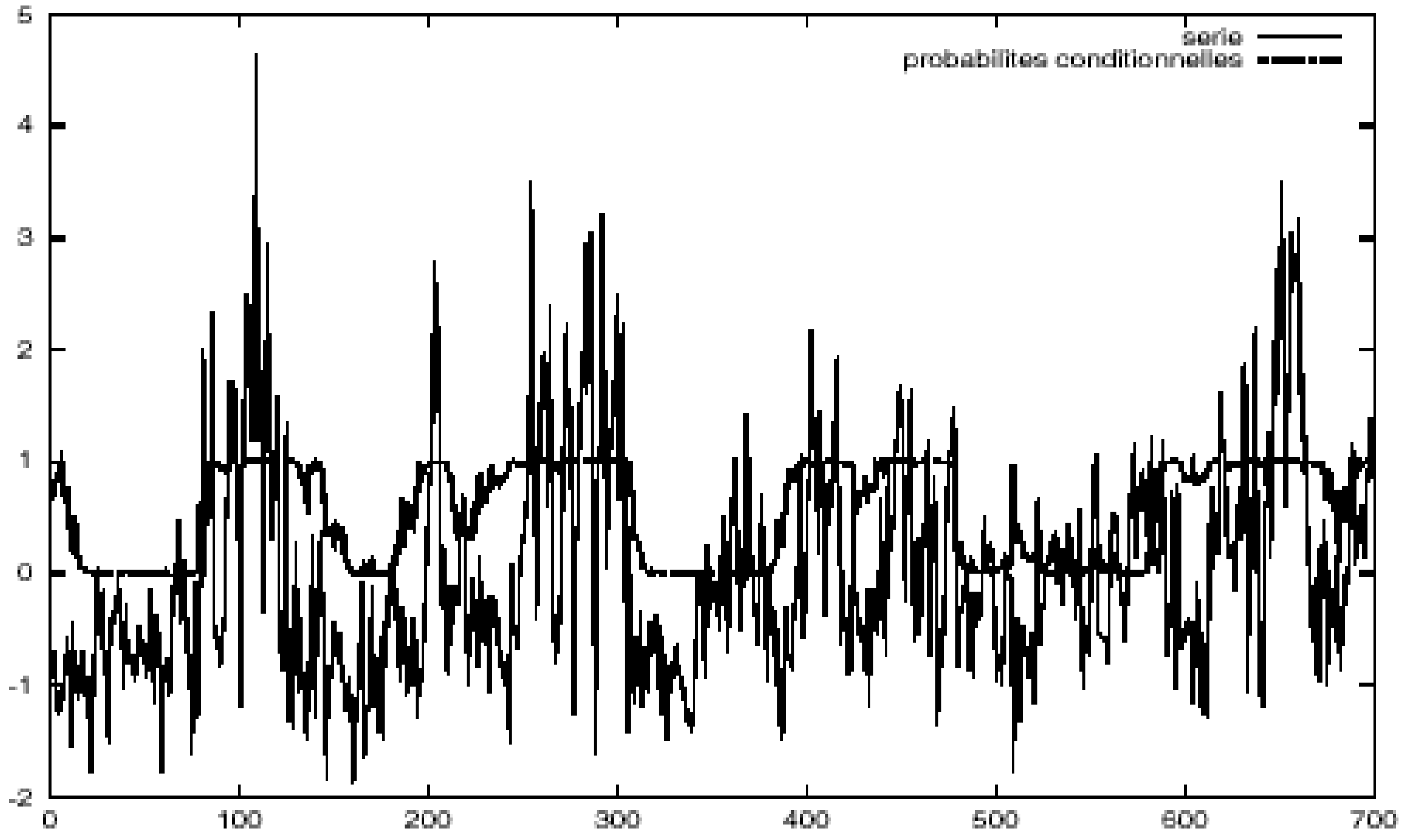
$$\begin{cases} \sigma_1 = 0.11 \\ \sigma_2 = 0.20 \end{cases}$$

The standard deviation of both models

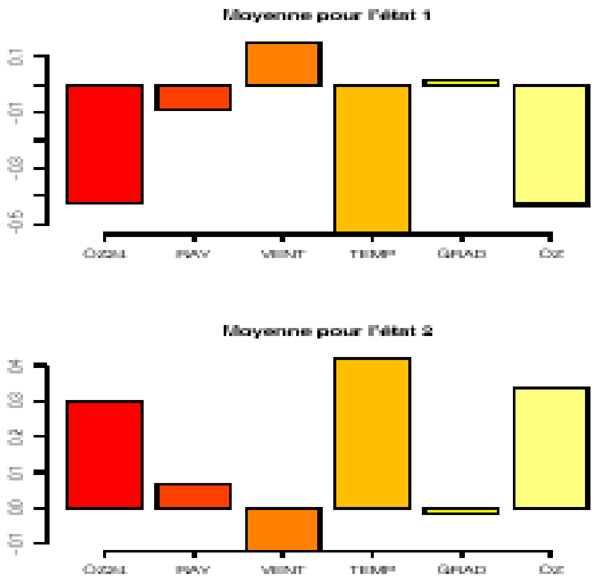
Quadratic error in sample and out of sample

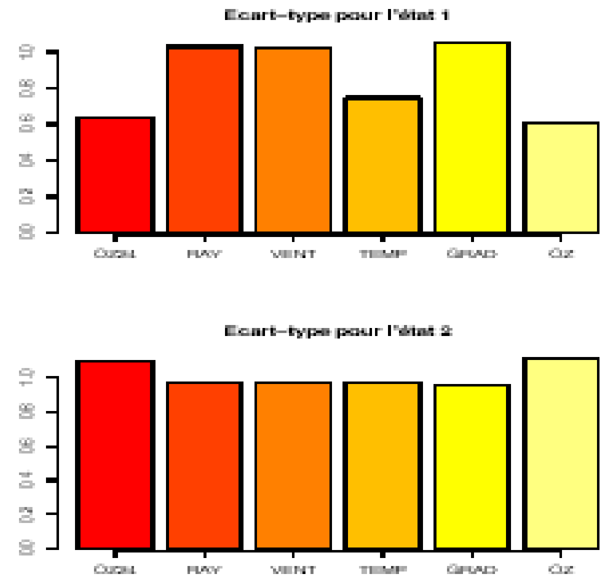
Années	1994-1996	1997
RMSE	16.51$\mu g/m^3$	16.75 $\mu g/m^3$

Probability to be in the state 2 (high values)



Mean and standard deviation of the variables in states 1 and 2

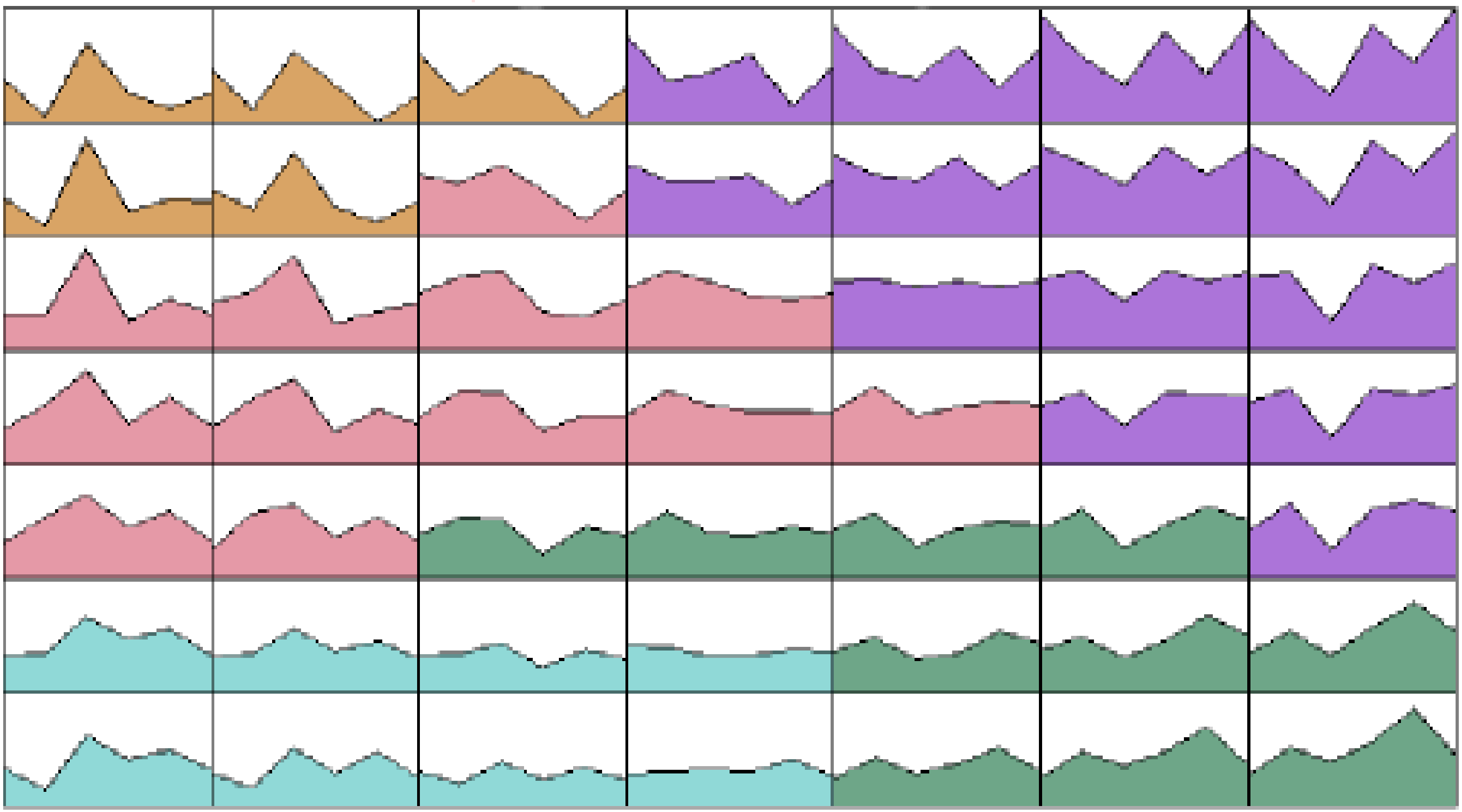




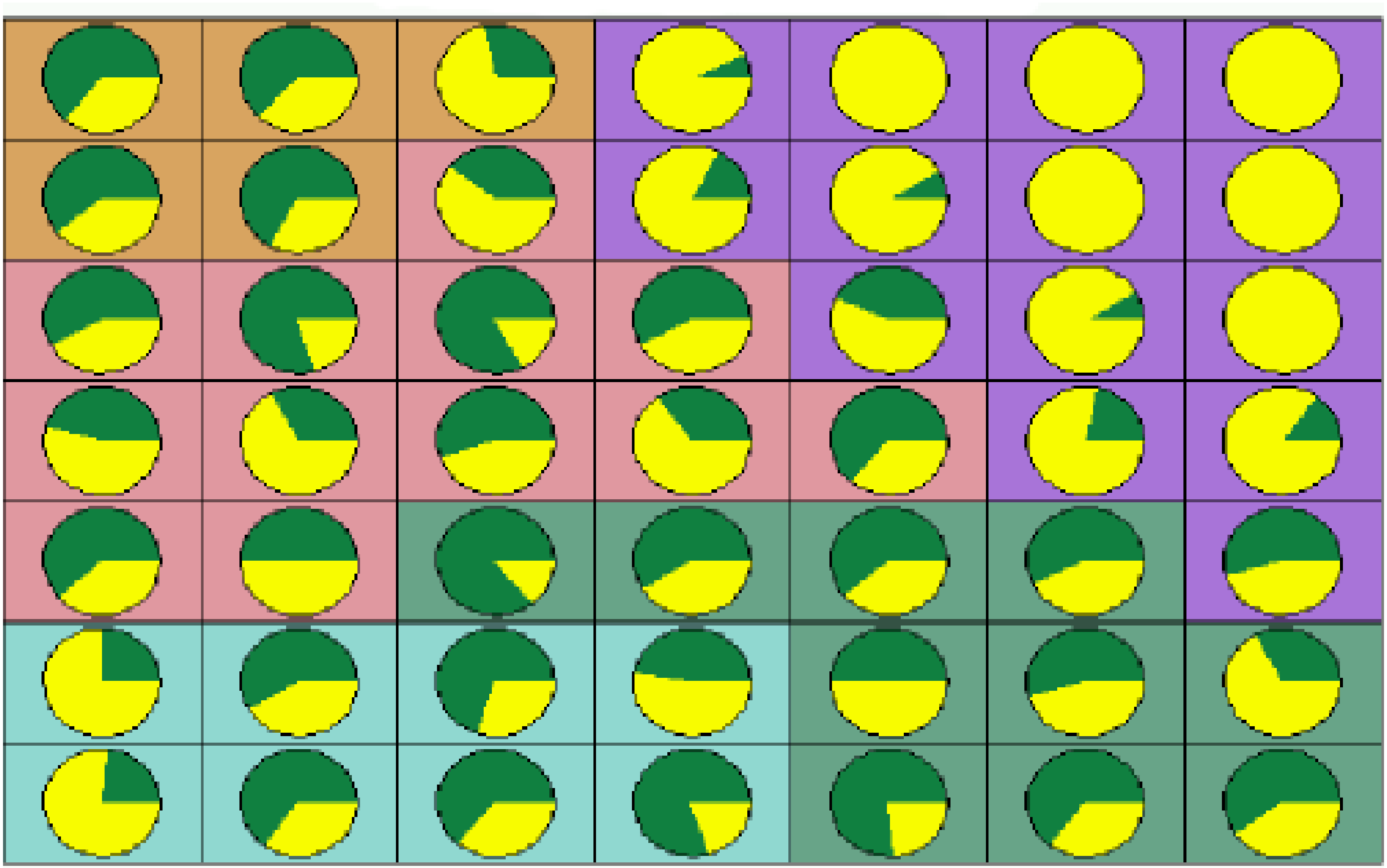
Means

Standard deviation

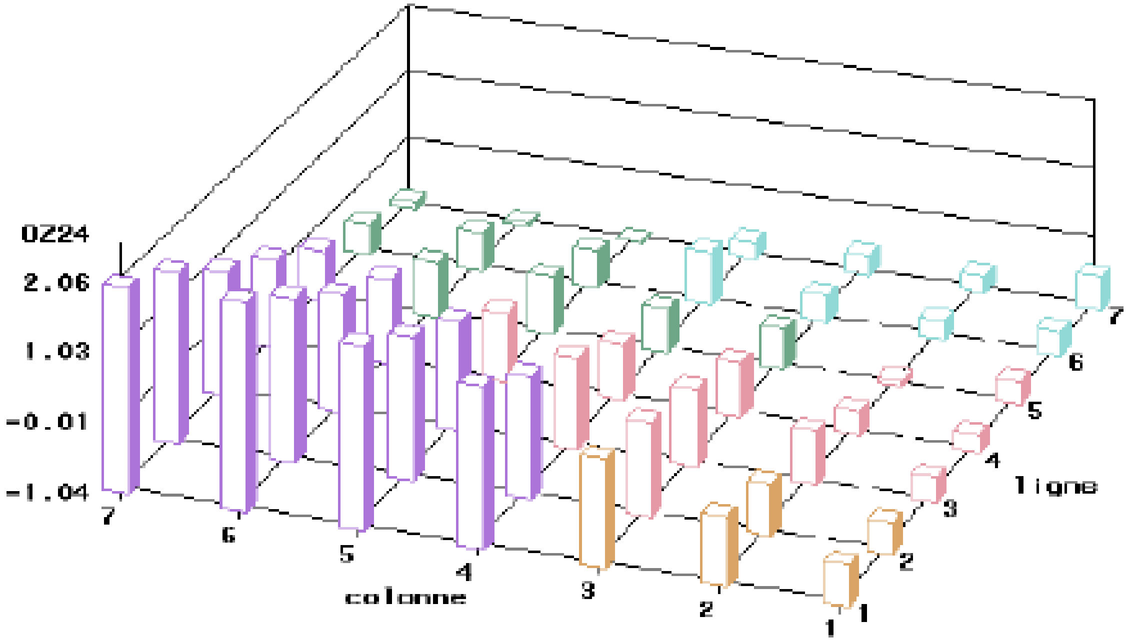
Kohonen map, and the 5 classes



Classes and super-classes, probability to be in state 1 or 2 (yellow for 2)



The ozone level 24 hours before (OZ24)



Interpretation

- The upper right corner contains the situations with high pollution levels, low wind, high temperature and gradient. Almost all the observations in this zone were identified by the non linear model, that is the state 2 of the HMM. Below, there are classes with observations whose values are near the means (except the temperature).

- The upper left corner contains the observations with low speed of wind and low gradient, etc. We can observe that the meteorological variables are not very discriminating to separate the hidden state 1 from the hidden state 2, which occurs in almost all the regions on the map, except the upper right corner which is specialized in the state 2.

Conclusion

- The Kohonen map is used to explain one partition of the data.
- We show that the meteorological conditions are not decisive
- In fact, it is necessary to add some components (past values)